

Moral Responsibility and Ignorance

Milo Nanni

Thesis submitted to the University of Nottingham for the degree of
Doctor of Philosophy

September 2017

Acknowledgements

I am very grateful to a number of people for supporting me in the last four years. Christopher Woodard, Penelope Mackie, and Jules Holroyd supervised my work and were extremely helpful with their thorough comments. Andrea Raimondi directed my attention to the paper I discuss in section 2.0 of chapter 2. Shinya Aoyama patiently listened to the problems I had, still unresolved, at the end of each working day. Nomy Arpaly's brilliant talks reminded me of how much we can be inspired by someone with whom we disagree. Zachary Hoskins made me feel welcome every time I knocked on his office door for advice. Mariateresa Tassinari offered me her invaluable philosophical counselling and moral support.

The Art and Humanities Research Council granted me with a three-year full tuition fees scholarship.

My greatest debt is with Andrzej, my partner, for his unique support through the good and bad times.

This thesis is dedicated to my great grandaunt, Rita Viroli, without whom it wouldn't have been possible for me to move to England.

Contents

Introduction.....	5
Chapter 1 – Attributionism and Volitionism.....	13
1.0.....	14
2.0.....	17
2.1.....	24
3.0.....	32
4.0.....	42
Chapter 2 – Culpability and Belief.....	44
1.0.....	46
1.1.....	54
2.0.....	62
3.0.....	71
Chapter 3 – Blameworthiness and Wrongness.....	73
1.0.....	75
1.1.....	80
1.2.....	83
2.0.....	91
2.1.....	95
3.0.....	103

Chapter 4 – Transferring Blame: Factual and Moral Ignorance.....	106
1.0.....	107
1.1.....	112
1.2.....	114
2.0.....	120
2.1.....	124
2.2.....	129
3.0.....	133
Chapter 5 – Culpability and Akrasia.....	135
1.0.....	136
1.1.....	141
1.2.....	150
2.0.....	154
3.0.....	163
Chapter 6 – Moral Responsibility and Consciousness.....	165
1.0.....	166
2.0.....	174
2.1.....	180
2.2.....	184
3.0.....	187
Conclusion.....	189
<i>Bibliography</i>	191

Introduction

The aim of this thesis is to defend a version of volitionism from objections concerning the epistemic condition of moral responsibility (especially of moral culpability). Thus, the purpose of my thesis will be mainly defensive, as I will not provide a positive argument for my volitionist view.

An agent is morally responsible in a positive sense when it is appropriate to praise her, while she is morally responsible in a negative sense when it is appropriate to blame her. Ever since Aristotle, many philosophers have agreed that moral responsibility has both a freedom condition and an epistemic condition. This thesis will focus on the epistemic condition, which concerns how the knowledge the agent either possesses or lacks affects her responsibility. For instance, the ignorance of the agent may sometimes undermine her responsibility.

Volitionism is a kind of account of moral responsibility according to which an agent is responsible only for what she has chosen, either directly or indirectly. For example, an agent can be responsible for an action she has directly chosen to perform and also responsible for an event that has resulted from that action. Objections to volitionism mainly come from attributionists. Attributionist (or character-based) accounts of moral responsibility claim that an agent is responsible only for what is appropriately attributable to her. For instance, an agent can be responsible for an attitude that is attributable to her, because it reflects her character. Also, she can be responsible for an action because it reflects that attitude. For the attributionists it is not necessary that the agent has chosen that attitude for her to be responsible for it. For the volitionists it is not sufficient that the agent's action reflect her attitude for

her to be responsible for the action. Take the example of laziness. Suppose I must decide whether to get up in the morning or to lie in bed. I am lazy and I lie in bed. For the attributionist I am responsible for acting lazily because my action reflects my laziness. For the volitionist I am responsible for acting lazily because I have chosen to succumb to my laziness.

Before proceeding with the summary of the thesis by chapter, I will present the view to be defended and I will clarify some of the central terms. My view states that an agent is morally blameworthy for her action only if (a) the action is morally wrong and (b) she has performed the action against her better judgement that the action is wrong or from a state of culpable ignorance.

First, note that I will use the terms 'blameworthiness' and 'culpability' as synonyms. This means that when I say that an agent is culpable, I mean that she deserves blame. Secondly, note that I am not committed to the claim that we should blame the agent every time she deserves blame. In order to clarify this, let me say something about how I understand blame. One way to explain my notion of blame is to imagine a scenario in which at the end of our life a supernatural judge gets to punish or reward us for how we have behaved. Suppose that this judge knows all of our actions and mental states. In addition, he knows which of our actions were morally right, wrong or merely permissible. If the majority of our actions were blameworthy the supernatural judge will punish us. This thought experiment should give the reader an idea of how I understand blame. In short, in this thesis I am interested in the metaphysical nature of blame independently of the practice of blame. In other words, according to my account, we may decide that we should not

blame a blameworthy agent or that we should blame a blameless agent. This might depend, for instance, on whether we think that blaming the agent could change her behaviour in the future. In this respect, blameworthiness is then neither practical nor forward-looking. It is instead backward-looking and it concerns the metaphysical status of the agent when certain conditions are satisfied.

Now I will clarify the two necessary conditions of blameworthiness for which I will argue: (a) and (b). Consider condition (a). In claiming that in order to be blameworthy an action must be morally wrong, I am employing a non-perspectival notion of wrongness. I take this to be the notion of wrongness most commonly assumed in the literature on blame. Non-perspectival wrongness stands somehow in between the notions of objective and subjective wrongness. While objective wrongness is independent of the circumstances and mental states of the agent, subjective wrongness depends on the belief of the agent about the moral status of the action. Now, whether an action is wrong in a non-perspectival sense does depend on the circumstances of the agent, but it does not depend on the belief the agent has about the moral status of her action. For instance, whether my driving a car is non-perspectivally wrong could depend on whether I have a driving license, but it is independent of whether I think that driving that car is wrong.

Let me now clarify condition (b). According to my account, to say that in order to be blameworthy an agent must act against her better judgement that her action is wrong means that she must have the conscious belief, at some point before the action, that she should not perform that action. In addition, the agent must act despite that belief (*e.g.* because she cannot be bothered to do the right thing). Alternatively, if the agent is ignorant of the wrongness of her action, she could still

be blameworthy for it, but only if that ignorance is culpable. I will argue that the agent's ignorance is culpable when it is caused by an earlier action which was performed knowingly, against the agent's better judgement.

Finally, I would like to outline some motivations for my view. The originality of my contribution concerns the conjoining of conditions (a) and (b). I will argue that the reason why we should not leave out condition (a) is that, from a moral point of view, we have the intuition that morally right actions should take place in the world as well as we have no moral reason to oppose to the occurrence of morally neutral actions. Therefore, it seems unreasonable to claim that an agent deserves blame for an action that is not wrong. After all, in performing an action that is not wrong, the agent does not bring about something which morally worsens our world. Now consider condition (b). The reason why we should not leave out this condition is that, as I shall argue, if an agent acts according to her better judgement that her action is right but she is simply mistaken and her action is in fact wrong, it seems unfair to claim that she deserves to be blamed. Because of our limitations as moral agents, when we aim at doing the right thing, we can only base our decision upon our better judgement, and whether our belief about the moral status of an action is mistaken is often out of our control (*e.g.* it may depend on our moral education). However, the agent's mistake could be itself culpable, hence the idea of culpable ignorance must be included in a necessary condition of blameworthiness. An agent may be indirectly culpable for an action performed from ignorance if that ignorance is itself the result of an earlier action performed against the agent's better judgement. If such earlier action was not performed, this means that the agent had no control over the mistaken belief she would form and over the resulting wrong action. And

claiming that in this case the agent deserves blame would go against our intuition that it is out of place to feel resentment towards an agent for something over which she has neither direct nor indirect control.

Let me now outline the structure of the thesis. In chapter 1 **I will provide reason in favour of volitionism and against attributionism to motivate further articulation of volitionism.** One reason I find volitionism superior to attributionism is that volitionist accounts are able to accurately distinguish between cases where the agent is blameworthy and cases where the agent is simply morally flawed. Attributionism tends to conflate these cases and therefore suggests we deserve blame more often than volitionists allow. Volitionist accounts claim, for instance, that while a bad attitude that reflects the agent's character might make her a *bad* moral agent, the attitude does not, by itself, make her a *blameworthy* agent. Attributionists cannot draw a correct distinction, because they do not accept a control condition on blameworthiness which is also inclusive of a strong epistemic condition.

Moral responsibility is most commonly attributed to agents for their actions. Actions are influenced by our beliefs. Thus, it seems appropriate to investigate whether we can be responsible for our beliefs in the first place. One difference between actions and beliefs is that we do not seem to choose to form a belief the same way we choose to perform an action. Therefore, beliefs seem to require a different account of responsibility from the account of responsibility for actions. In chapter 2 I will discuss whether and when it is appropriate to blame an agent for holding a false belief. Volitionism claims that blaming someone for holding a false belief is appropriate only if she has direct or indirect voluntary control over that

belief. By contrast, attributionism claims that holding a false belief may be blameworthy regardless of whether we have voluntary control over it, if holding that belief is an epistemic vice which reveals a defect of our character. **I will argue that it is appropriate to blame an agent for a false belief only if she chose not to inform herself and she knew she should, in order to avoid forming a false belief.**

Although beliefs can sometimes be blameworthy, we most commonly blame for wrongful actions. Some philosophers have reflected on whether it is appropriate to blame for actions that are not morally wrong. In chapter 3 **I will defend from some objections the thesis that an agent is blameworthy for performing an action only if the action is objectively wrong (the Objective View).** The philosophers who raise objections to the objective view claim that it is not necessary that an action be objectively wrong for the agent to be blameworthy for it. Their claim is rather that blameworthiness is affiliated only with the agent's mental states at the time of action.

Blameworthiness requires wrongness but not all wrongful actions are culpable actions. For instance, a wrongful action done from ignorance may be blameless. Some philosophers have discussed under which conditions this is the case. In chapter 4 **I will defend from some objections the thesis that whenever an agent acts from ignorance, she is culpable for the act only if she is culpable for the ignorance from which she acts (the Ignorance Thesis).** Volitionists support this thesis and claim that ignorance is culpable when it is the result of an earlier culpable action. Objections to the Ignorance Thesis mainly come from attributionists, who hold instead that in order to explain responsibility for ignorant behaviour, we do not need to trace back to culpable actions that caused that ignorance. I will show that the

Ignorance Thesis applies both to cases of factual ignorance and to cases of moral ignorance.

If the Ignorance Thesis is true, one may then ask what makes an agent's ignorance culpable. Attributionists claim that ignorance is culpable if it reflects a bad character trait of the agent (*e.g.* lack of concern). Volitionists claim instead that ignorance is culpable if it is the result of an akratic action. An agent acts akratically when she performs an action believing she should not perform it. Volitionists claim that an agent is culpable for an action only if that action is either akratic or culpably ignorant. Attributionists, on the other hand, reject this claim, since for them an action that is *not* the result of an akratic choice may still be blameworthy, provided that it reflects the agent's character. In chapter 5 **I will defend from some objections the thesis that moral culpability always involves akrasia (the Akrasia Thesis).**

Finally, if the Akrasia Thesis is true and an agent must act against her better judgement in order to be culpable, then culpability for actions requires that the agent have some awareness of her action. There is a debate among volitionists regarding what type of awareness is required for moral responsibility and what it is exactly that the agent must be aware of. Some believe that dispositional awareness is sufficient for moral responsibility, while others claim that occurrent awareness is the type of consciousness that grounds responsibility. Furthermore, some believe that we must be aware of the facts that make our actions good or bad in order to be responsible for them, while others think that we must have the conscious belief that our action is right or wrong. In chapter 6, I will conclude the defence of my version of volitionism. First, **I will undermine the thesis that in order for an agent to be morally responsible for an action, it is necessary (and sufficient when the other**

conditions are met) that some facts she takes to be playing a role in explaining why the action is good or bad be personally available to her (Consciousness Thesis). Focussing on moral culpability in particular, I will advance my own formulation of the Consciousness Thesis, according to which in order for an agent to be blameworthy for an action she must occurrently token – at some point before the action – the belief that the action is wrong, or at least the belief that she should not perform the action.

Chapter 1 – **Moral Responsibility: Attributionism and Volitionism**

In his paper "The Good, The Bad and the Blameworthy", Neil Levy identifies two main kinds of account of moral responsibility. *Volitionism* claims that an agent is morally responsible for what she has chosen – either directly or indirectly. *Attributionism* claims that an agent is morally responsible for what is attributable to her. Both kinds of account agree that when we say that an agent is morally responsible we mean that certain attitudes towards her – and towards what she is responsible for – are appropriate, especially praise and blame.¹

The main purpose of this chapter is to explain some differences between these two kinds of account and to give sufficient reason in favour of volitionism and against attributionism to motivate further articulation of volitionism. I will agree with Levy that one of the main flaws of the attributionist account is that it fails to take into consideration a stronger epistemic condition on moral responsibility (an epistemic condition which will be the focus of the following chapters).

In 1.0 I will introduce the distinction between attributionism and volitionism via Watson's distinction between aretaic and accountability faces of responsibility, which Watson sees as irreducible kinds of responsibility. Attributionists believe that the accountability face of responsibility is not genuine moral responsibility, whereas the volitionists believe that it is instead the aretaic face of responsibility that does not express genuine moral responsibility. In 2.0 and 2.1 I will present two examples of attributionist account defended, respectively, by T.M. Scanlon and Angela Smith. Finally, in 3.0, I will outline Levy's considerations against attributionism and in

favour of volitionism, as well as presenting his specific objections to claims made by Scanlon, Smith and Watson.

1.0

In this section I will illustrate Gary Watson's distinction between what he believes to be two kinds of responsibility: the *aretaic* and the *accountability* faces of responsibility. The aretaic face of responsibility seems to be what the attributionists see as genuine moral responsibility. The accountability face of responsibility appears to be what the volitionists instead call moral responsibility. According to most attributionists and volitionists the kind of responsibility defended by their opponents is not genuine moral responsibility.²

In his paper "Two Faces of Responsibility", Gary Watson claims that there is one conception of responsibility that is connected to a view of free and autonomous action. According to this conception, to be responsible is to possess the power to "put various ends before the self" and "to be governed in action by the thought of some end to be reached". This is what Dewey calls 'moral capacity', which entails "freedom from the appetites and desires" and can be seen as "the power of self-government".³ This view, which Watson calls the *self-disclosure* view, expresses the aretaic face of responsibility. Watson claims that this view is problematic, and that is why we need to take into consideration also an accountability aspect of responsibility. According to Watson, the self-disclosure view does not say anything about the content of the agent's ends, about their ability to understand the grounds of

² In using Watson's distinction as a way of introducing the attributionist/volitionist distinction I am following Levy (2005).

³ Dewey (1891); reprint, New York: Hillary House, 1957, pp. 160-161.

moral requirements, or about the authority to hold responsible agents to these requirements.⁴ According to the abstract characterization of the agent's self-government, even an insane or morally dead agent can be a self-governing agent. Therefore, this power of self-government cannot explain our practice of holding one another morally accountable. This is why according to Watson responsibility has also another face which is accountability.

According to the self-disclosure view, the capacities identified in the agent make her conduct attributable to her and make her, as an adopter of ends, open to appraisal. This attributability is what Watson identifies as the aretaic face of responsibility. This kind of responsibility is about the quality and the character of the agent's life.⁵ According to Watson, the aretaic conception indeed expresses a genuine kind of responsibility, but it does not capture the accountability aspect of responsibility. The aretaic appraisal is moral in the sense that it is concerned with what are the most choiceworthy things in life.

If someone betrays her ideals by choosing a dull but secure occupation in favor of⁶ a riskier but potentially more enriching one, or endangers something of deep importance to her life for trivial ends (by sleeping too little and drinking too much before important performances, for example), then she has acted badly – cowardly, self-indulgently, at least unwisely.⁷

On the other hand, Watson notes that in order for this person to be held responsible in the accountability sense, we must think that she is accountable to us or to someone else, rather than her behaviour being simply her own business, and

4 Watson (2004b), pp. 263-264.

5 *Ibid.*, pp. 264-265.

6 Here I take 'in favor of' to mean 'in preference to'.

7 Watson (2004b), pp. 267-268.

nobody else's.⁸

According to Susan Wolf the responsibility understood by the self-disclosure view is about the relationship between the agent's conduct and her deepest values or commitments – her “real self”. The relation between the agent's real self and her behaviour is executive and expressive. We evaluate the agent because her conduct expresses her evaluative commitments and the adoption of some ends. Watson claims that adopting some ends is equivalent to declaring what we stand for. When my actions flow from my values and ends, they express my agential identity. Therefore, there is ethical depth in aretaic appraisals, and committing ourselves to a conception of value can be seen as taking responsibility for it, because when we stand for something we are also ready to answer for it.⁹

From the aretaic perspective, control affects responsibility only if its absence makes the agent's conduct not attributable to her. Watson points out that one issue for the self-disclosure view is to explain what it means to adopt some end, since brain-washing and hypnosis can generate motivation in the agent without her being responsible for it. This view must then explain how attributability is undercut by these processes.¹⁰ On the other hand, when we hold someone accountable there is an issue of whether it is fair to do so and, because of this concern, control is a condition of moral accountability, while it does not apply to the aretaic appraisal.¹¹ It seems to me that what Watson means here is that there is no fairness involved when the action is simply not attributable to the agent, because we want to say that the attribution of the action (that is, of responsibility, in the aretaic sense) would be simply mistaken,

8 *Ibid.*, p. 268.

9 *Ibid.*, pp. 271-272.

10 *Ibid.*, p. 273.

11 *Ibid.*, pp. 274-275.

rather than unfair. On the other hand, in the case of accountability, two actions could be both appropriately attributed to the agent, but for one it might be unfair to hold her accountable if she lacks control over it (a kind of control that does not undercut the attribution of the action to the agent).

The aretaic and the accountability perspectives involve two different approaches to moral blame. Blaming in the aretaic sense is judging the conduct of the agent as faulty, whereas in terms of accountability blaming is rather a response to those faults. While from the aretaic perspective blame is deserved as long as the attribution of fault is warranted, in terms of accountability blame is deserved when it would be an appropriate unfavourable attitude toward the agent for his conduct.¹²

Finally, Watson maintains that an agent who is a subject of aretaic appraisal must be able to apply moral predicates to her deliberations. This agent must be intelligent and sensitive enough to understand the normative concepts in terms of which the appraisal is conceived. For example, an agent cannot be judged as cruel if she does not possess the concept of cruelty.¹³

2.0

Levy notes that, although the attributionist view seems to reflect the aretaic perspective while the volitionist view appears to reflect the accountability perspective, most attributionists and volitionists disagree with Watson that these are simply two irreducible kinds of responsibility. They believe (and I agree with this)

¹² *Ibid.*, p. 279.

¹³ *Ibid.*, p. 283. Note that the claim that in order for an agent to be judged (in the aretaic sense) for being cruel she must possess the concept of cruelty is stronger than the claim that she must be intelligent and sensitive enough to understand the concept of cruelty. Moreover, Watson's claim seems plausible as far as cruelty is concerned, but it might not apply to other character traits (e.g. impulsivity).

that the kind of responsibility defined by their opponents is not genuine responsibility.¹⁴ Thus, it seems that we have to choose between the attributionist and the volitionist accounts and not simply adopt both of them and claim that they represent different aspects of responsibility. One reason for this is that both attributionists and volitionists believe that the appropriateness of reactive attitudes, such as blame and praise, is relevant when we hold an agent responsible, and there seems to be some agreement on (or it is at least less controversial) how these reactive attitudes manifest themselves.¹⁵

In order to discuss more in detail some of the differences between attributionism and volitionism, I will start with presenting, in this section, an example of attributionist position, by illustrating Scanlon's view of responsibility. I will later consider objections made to some of Scanlon's claims by Levy, who is a volitionist.¹⁶ Scanlon believes that moral criticism is appropriate when a faulty process of self-governance can be attributed to the agent, that is, when there is a connection between the agent's bad action and her faulty judgment-sensitive attitudes. According to the volitionist such a condition is too weak, as it does not include a condition of choice that must be satisfied: the agent must choose her action. Thus, it is not sufficient that the action is connected to her faulty judgment-sensitive attitudes for the agent to be morally criticised for it, according to the volitionist.

T.M. Scanlon, in his book *What We Owe to Each Other*, clearly states that "questions of 'moral responsibility' are most often questions about whether some action can be attributed to an agent in the way that is required in order for it to be a basis for

14 Levy (2005), pp. 3-4.

15 The disagreement, if anything, might then be about what it means for a reactive attitude to be appropriate. I will not discuss this issue here.

16 See Levy's objections in section 3.0.

moral appraisal".¹⁷ For Scanlon moral reasoning is our attempt to find the principles that we can all be asked to employ in deliberation. Those principles are the basis for moral criticism. According to Scanlon, we morally criticise an agent when her way of governing herself is not allowed by any reasonably unobjectionable principle. Moral criticism calls for the agent to explain why the moral criticism is mistaken or to acknowledge that her self-governance was indeed faulty.¹⁸

Scanlon maintains that what makes an action blameworthy is that the agent either did not take into account or knowingly acted against what should have counted as a reason against her action. And what makes it appropriate for a person to feel resentment towards the agent is that the requirements flowing from the person to whom justification is owed have been ignored by the agent, because of her faulty self-governance. The agent has violated requirements of justifiability and hence others can appropriately react with indignation. If the agent takes a value seriously but still acts contrary to a reason that flows from that value, then the agent is open to criticism coming from other people who share that value. According to Scanlon, moral criticism applies to those capable of reflective self-governance in regard to the attitudes that should be under the control of reason. By criticising (morally) we claim that the agent has governed herself in a way that cannot be justified, according to morality, and that she should acknowledge this and apologise (or, otherwise, explain why the criticism is inappropriate). Moral criticism for an action is appropriate even if the agent has performed the action inadvertently, if advertence is called for in that situation.¹⁹

17 Scanlon (1998), p. 248. Scanlon also accepts the existence of other kinds of responsibility such as 'substantive responsibility'. I do not discuss this here as it does not concern the debate on which I focus.

18 *Ibid.*, p. 268.

19 *Ibid.*, pp. 271-272.

Thomas Nagel, in his paper "Moral Luck", cites cases of people who have vices but act perfectly because of their effort of will.²⁰ Scanlon notes that a greedy person, for example, might indeed struggle against her tendencies. She might think that getting more (*e.g.* power) is a good reason to act in a way that allows her to do so, when that is in fact not a good reason. Even if she affirms the opposite judgment when this tendency occurs, we can still attribute it to her. After all she herself sees it as hers, and it would be appropriate to ask her to modify it. Thus, she should still be judged morally for that tendency, because if she did not have it, she would be a better person.²¹

According to Scanlon, moral criticism is about the judgment-sensitive attitudes of the agent. The person who is morally criticised is asked to explain why the criticism is not justified or, if it is, she is asked to withdraw or modify those attitudes. However, Scanlon points out, I might appropriately criticise someone even if her attitudes could not be modified, since what matters is that the attitude be under the control of reason in an ideally rational agent (for that is what 'judgment-sensitive' means). In some cases criticising someone may cause so much pain that it would be wrong to express that moral criticism. This does not mean that the criticism is not appropriate.²²

Scanlon identifies three categories of conditions under which we should modify or withdraw a moral appraisal. First, there are conditions that show that the action is not attributable to the agent.

(...) it would be inappropriate to credit or blame me for a particularly offensive remark if it

20 Nagel (1979).
 21 Scanlon (1998), pp. 273-274.
 22 *Ibid.*, pp. 275-276.

turned out that the remark was produced by someone else, who caused me to utter it by stimulating my brain with electrodes. The same would be true if the remark were produced by posthypnotic suggestion or uttered while sleepwalking (...) [or] in the case of behaviour induced by drugs or by sudden episodes of mental illness.²³

In these cases, Scanlon notes, there is no connection between the action and the agent's judgment-sensitive attitudes. Hence, asking the agent for a justification is pointless. If the act does not follow from the agent's judgment-sensitive attitudes, it is not *her* act in the sense relevant to moral appraisal.

Suppose (...) that someone who has previously always been kind and considerate suddenly begins making cruel and wounding remarks to her friends after being hit on the head or given drugs for some medical condition (...). (...) after fifteen years, the person still behaves in this way and shows no sign of rejecting these attitudes or finding them 'alien' to her (...).²⁴

Scanlon claims that since in this case the event (the accident or the drug consumption) has marked a change in what the agent is like, it is likely that, after fifteen years, our appraisal of the agent will shift, as we might properly criticise the agent for her new attitudes.

The second category of excusing conditions discussed by Scanlon includes ignorance and mistakes of fact, as conditions that alter the action's character.

I may think, for example, that your stomping on my foot reflects a lack of concern for my pain and for my right to be free from unwanted touching. But this appraisal of your action

23

Ibid., pp. 277-278

24

Ibid., pp. 278-279.

must be modified if I learn that you assumed that the toy spider on my boot was real and, having had experience in the tropics, you thought that you were saving my life by killing it before it had a chance to bite me.²⁵

This is a case, Scanlon claims, in which the action is attributable to the agent but it is an action that does not indicate a blameworthy attitude, as it was not ill-intended.

Finally, according to Scanlon, the conditions belonging to the third category are those under which the person does not possess the capacities that are presupposed by moral agency. When an agent is not able to assess reasons, or when her actions are not affected by her judgments, because of mental illness or defect, she does not take part in the system of co-deliberation that is moral criticism.²⁶

Then Scanlon considers the following example offered by Susan Wolf:

(...) [Suppose that] a woman fails to give her friend a book that she very much wants because, as a result of her "personality and social development", she is either "too self-centered for the thought, 'My friend would like this book' to occur to her" or "so unfamiliar with examples of sincere, non-instrumental friendships that the thought 'I should buy this book, just to make my friend happy' cannot help appearing irrational to her".²⁷

Scanlon maintains that the case in which the woman is too self-centred is a case in which pleasing her friend is not something important to her. Hence, moral criticism is warranted, as not placing value on other people's interests gives a strong reason to see that person as bad and as a wrongdoer. And for such a person moral

25 *Ibid.*, p. 279.

26 *Ibid.*, p. 280.

27 *Ibid.*, p. 283.

criticism is appropriate.²⁸ For Scanlon, if the judgment that a person has acted wrongly is appropriate, then morally criticising her is appropriate.²⁹ We can appropriately blame someone for an action if the action is wrong and attributable to the agent. In other words, if the faulty process of self-governance can be attributed to the agent, moral criticism of that agent is appropriate.³⁰

Scanlon points out that, unlike moral criticism, evaluations of physical attributes do not concern judgment-sensitive attitudes and they do not ask the person possessing those physical attributes to revise them. Moreover, unlike other types of criticism of judgment-sensitive attitudes, moral criticism has a particular significance for the relations the agent has with other people.³¹

In Scanlon's view, the fact that an agent cannot appreciate the force of a moral reason does not make her immune from moral criticism for not giving that reason appropriate weight. Even if she cannot see how that reason counts against her judgment-sensitive attitude, the attitude may be still attributable to her as well as the errors involved in it. If so, it would still be appropriate to morally criticise that agent for holding that attitude.³²

As a final remark, Scanlon clarifies that the kind of unfreedom that undermines his notion of responsibility is only the unfreedom which undermines attributability: "(...) [Imagine a] 'willing addict' [who] prefers the life of addiction, wants to act on the desire to take the drug, and would do so whether or not he was able to resist".³³ According to Scanlon, the willing addict deserves moral criticism because his action reflects his assessment of the relevant reasons, whereas an

28 *Ibid.*, p. 284.

29 *Ibid.*, p. 285.

30 *Ibid.*, pp. 286-287.

31 *Ibid.*, p. 287.

32 *Ibid.*, p. 289.

33 *Ibid.*, p. 291.

unwilling addict would be unfree in a way that makes the action of taking drugs not attributable to him.

To summarise, according to Scanlon's attributionist account, moral criticism is appropriate when the faulty process of self-governance can be attributed to the agent, that is, when there is a connection between the agent's bad action and her faulty judgment-sensitive attitudes. This account differs from volitionism because for the agent to be appropriately morally criticised no condition of choice must be satisfied. That is, Scanlon's attributionism does not require that the agent choose the action that is connected to her faulty judgment-sensitive attitudes. For instance, according to Scanlon, moral criticism towards a greedy person is appropriate not because she has chosen not to act against her tendencies, but simply because those tendencies are attributable to her and it would be appropriate to ask her to modify them.

2.1

In this section I will illustrate Angela Smith's view of responsibility as presented in her paper "Responsibility for Attitudes: Activity and Passivity in Mental Life". Like Scanlon, she is an attributionist. According to her account, an agent is morally appraisable for her mental state if this mental state is *supposed to be* governed by the agent's evaluative judgments, and if it is, *in principle*, open to modification or revision through the agent's reflection.

Smith's account begins with the claim that we often respond to people's voluntary actions in the same way as to their reactions and attitudes, and we tend to

think that it is important to take into consideration involuntary responses in order to properly assess a person. Smith believes that involuntary responses significantly indicate who a person is, morally speaking, and what are her moral commitments, because these involuntary responses probably mark her genuineness and sincerity more than voluntary actions.³⁴

Smith points out that generally we do not deliberately notice or fail to notice things around us. We can only choose to pay attention to something and not to something else, but we cannot directly choose to notice something. Still, what a person notices or fails to notice may be taken as something having great *expressive significance*. For instance, if a member of my family is not sensitive to factors regarding my welfare, I may take this as an indication that they do not care about me as they should. Or if it never occurs to a friend of mine to do something to make me happy, it looks like some criticism from me is appropriate, because my friend does not seem to have my happiness as their end. Smith suggests that even if my friend replied that they do not have control over what occurs to them, this would not be a pertinent response. After all, having an end is also about finding certain things important, and even if my friend professed their concern for me, this would not sound like a genuine or deep concern, in light of their failure to appreciate factors bearing upon my welfare.³⁵

But even if we leave aside special relationships, Smith maintains that “If I do not notice when my music is too loud, when my advice is unwelcome, or when my assistance might be helpful to others, these can be described as ‘involuntary failings’ (...) [which] are commonly taken to be an appropriate basis for moral criticism”.³⁶

34 Smith (2005), pp. 241-242.

35 *Ibid.*, pp. 242-243.

36 *Ibid.*, p. 244.

Smith calls this a form of moral insensitivity which indicates that I do not take certain needs and interests seriously. After all, we tend to think that, if something is important to a person, it will have some rational influence on her tendency to notice factors regarding the welfare of that thing. So her failing to notice these factors indicates that she does not really accept the evaluative judgment that that something (or someone) is important to her.³⁷

Smith notes that what comes into thought is not under our direct voluntary control and considers the example of a “man who in the course of a discussion of how to deal with political or business rivals says, ‘Of course, we could have them killed, but we should lay that aside right from the beginning’”.³⁸ The reason why we feel uneasy with this person – Smith claims – is that we think that the fact that killing came to his mind as an option shows that he might not be really committed to the judgment that we must not kill to further our own interests. Even when he rejects that option we might still think that he is not the sort of person who places great value on human life.³⁹

According to Smith, the same applies to forgetting and desiring. If I have the sympathetic spontaneous desire to help others, it is not simply that it is fortunate that nature supplied that desire to me, but I deserve moral credit for it. This desire is explicable in the context of a more general evaluative judgment and commitment of mine to valuing others’ welfare, given that judgments and commitments have a rational influence on my desires. Smith thinks that the reason why we sometimes feel bad about forgetting something is that we realize that it can reasonably be taken to express a lack of appreciation for what we have forgotten. When we apologise for it,

37

Ibidem.

38

Smith (2005), p. 246.

39

Ibid., p. 247.

what we might be doing is simply vowing to care more in the future about what we have forgotten, hoping to strengthen our commitment to the relationship with the person to whom we are apologising.⁴⁰

Spontaneous emotions and reactive attitudes are also responses over which we do not exercise direct control – Smith notes. Still, we take things such as gratitude or envy to possess great expressive significance as they inform us about our deeper commitments and values. Smith considers cases where the agent’s action brings about harmful consequences which are unavoidable. In such cases, it might be that the agent’s commitment is not expressed by her action, and that is why she might feel regret for what she has done. If no regret were felt – Smith observes – we might suspect that the agent does not appreciate how significant is the harm she has contributed to cause, although she was blameless in doing so.⁴¹ And this suggests, according to Smith, that our reactions and our fundamental commitments and evaluative judgments are rationally connected or, in other words, that our attitudes are, at least partially, constituted by certain evaluative judgments: “To feel contempt toward some person, for example, involves the judgment that she has some feature or has behaved in some way which makes her unworthy of one’s respect”.⁴²

If we pursue projects of self-improvement we might influence our mental states and attitudes. Smith acknowledges this but believes it is not relevant to explaining the nature of the moral appraisals we make or to explain why we base our appraisals on those attitudes. After all, our appraisals are about the content of our attitudes and not about their origin. Smith thinks it is a mistake to attribute responsibility for holding certain attitudes in terms of their connection to prior or

40 *Ibid.*, pp. 247-248.

41 *Ibid.*, pp. 249-250.

42 *Ibid.*, p. 250.

future voluntary actions of the agent who displays those attitudes. Doing that prevents us from seeing the relation we have with our attitudes, that is, the fact that they reflect what we judge as valuable. Smith maintains that reasons and justifications can be requested from the person who holds an attitude because the attitude is rationally connected to her evaluative judgments, that is connected to her deeper tendencies to judge certain things as significant in her view of the world. These are judgments that need not be consciously recognized by the agent holding them, but we might find out about them when we have certain responses.⁴³

I may not realize, until I am faced with a choice, that I value the intellectual freedom and autonomy associated with a career in academia more highly than the economic rewards and benefits associated with a career in law. Or I may discover in some situation that I care more about being liked by others than I do about standing up for my moral principles.⁴⁴

Smith claims that when we realize that we hold a certain evaluative judgment, we can modify or withdraw it if we think there is no adequate justification to hold it. According to Smith, the way mental states and evaluative judgments are connected is that, rationally, a mental state *should* occur (or not occur) if I hold the relevant evaluative judgment.

(...) If I sincerely judge that there is nothing dangerous or threatening about spiders, I should not be fearful of them. The emotion of fear is conceptually linked to the judgment that the thing feared is in some way dangerous or threatening; (...) Occasionally, however, this connection can fail: I may continue to fear spiders even though I judge them not to be

43

Ibid., pp. 251-252.

44

Ibid., p. 252.

dangerous or threatening in any way.⁴⁵

Smith notes that our fear is normally justified by what we judge to be dangerous or threatening. If we stop thinking that what we fear is dangerous or threatening, we normally also stop being fearful. But if I continue to be fearful, I could be judged as *irrational*. So the fact that our mental states are not always sensitive to our evaluative judgments simply shows that we are not always rational, and that we deserve some form of rational criticism. In Smith's view, what matters is that my mental states are 'in principle'⁴⁶ sensitive to my evaluative judgments. Rational criticism is applicable also when a person is mistaken in her evaluative judgments. If I fear a spider because of my judgment that a particular spider is poisonous when it is not, my fear is due to my judgment being mistaken. If I disrespect a class of people because my evaluative judgment is that they are intellectually inferior because of their race, I can be appropriately criticized both rationally and morally.⁴⁷

Smith points out that, in cases of irrationality, there might be a conflict between the attitude that indicates that the agent has a certain evaluative judgment and the agent's sincere explicit claim that denies that judgment. Thus, the attitude in this case does not tell the whole story about the agent's evaluative judgments. We might then need to look at other explicit beliefs the agent holds in order to correctly infer from her attitudes to the judgments she accepts and check what is in fact attributable to her. When the connections among the beliefs and judgments the agent

45

Ibid., p. 253.

46

The expression 'in principle' is ambiguous in Smith's account. Levy highlights it as a weakness of the account as I will illustrate in section 3.0.

47

Smith (2005), pp. 253-254.

seems to accept break down, we talk about irrationality.⁴⁸

Smith then claims that in order for a mental state to be attributable to an agent so that she could be morally appraised, it must be a mental state that is, at least, *supposed to be* governed by (*i.e.* normatively connected to) the agent's evaluative judgments, and, in principle, open to modification or revision through the agent's reflection.⁴⁹ Thus, among the mental states included in this definition there are ordinary beliefs, desires, emotions, reactive attitudes and moral perceptions. Of course, also our deliberate choices and actions are, or should be, sensitive to our evaluative judgments, and are therefore, in the same way, subject to moral appraisal.⁵⁰

Smith claims that the conditions under which we form our evaluative judgments might be relevant to moral appraisal.

Consider (...) Abigail, who is raised in a family or community that is deeply racist, say, or religiously intolerant (...) [and] develop[s] evaluative tendencies and corresponding attitudes in line with those she sees operative in her family or surrounding community. As an adult, her attitudes (...) continue to reflect the vicious evaluative judgments thus formed in her childhood. Now contrast this person to another, Bert, who was raised in a loving and tolerant home and community, but who later in life reflectively comes to adopt racist and intolerant values.⁵¹

Smith believes that there is a difference in the moral assessment of Abigail and Bert in terms of the formation of their judgments. Unlike Bert, Abigail is not

48 *Ibid.*, pp. 254-255.

49 *Ibid.*, pp. 256-257.

50 *Ibid.*, pp. 262-263.

51 *Ibid.*, p. 267.

solely responsible for *developing* racist attitudes since they started developing before she could reflect on her evaluative judgments. However, in Smith's view, there is no difference between Abigail and Bert in terms of responsibility for *holding* racist and intolerant attitudes, because as long as those attitudes are sensitive to their evaluative judgments, they are attributable both to Abigail and to Bert. Since Abigail has now reached rational maturity and she can be asked to defend the evaluative judgment her attitudes reflect, in Smith's account she can be properly criticized for holding them.⁵²

In conclusion, Smith claims that holding someone responsible is recognizing that she fully participates in the moral community and in practices of interpersonal justification. It means that she is in principle able to regulate her conduct and attitudes and that it is appropriate to ask her to account for them.⁵³

To summarise, according to Smith's attributionist account, an agent is morally appraisable for her mental state if this mental state is *supposed to be* governed by the agent's evaluative judgments and if it is, *in principle*, open to modification or revision through the agent's reflection. This account differs from volitionism because it does not require that the agent chose to perform some earlier action knowing that it will bring about that mental state, or that she chose not to perform some earlier action that will prevent that mental state from occurring. For instance, Smith claims that involuntary failings, such as not noticing that my assistance might be helpful, are bases for moral criticism even if I did not choose to disregard others.

3.0

52

Ibid., pp. 267-268.

53

Ibid., p. 269.

In this section I will let the volitionist account emerge by drawing from Neil Levy's objections to attributionism. He intends to show that volitionism is superior to attributionism, a claim I agree with. Levy believes that attributability is not genuine responsibility, since reactive attitudes such as praise and blame are not appropriate just because we can justifiably attribute a certain attitude to the agent. Instead, moral responsibility requires stronger control conditions than those usually mentioned by the attributionists. Among those conditions, some important ones are epistemic, as I shall explain in what follows. An attitude can be attributed to an agent without any stronger control conditions being satisfied, but in this case reactive attitudes such as praise and blame will not be justified.⁵⁴

As I shall explain in this section, Levy argues for the superiority of volitionism by claiming that attributionists are mistaken in not drawing a distinction between blameless judgment-sensitive attitudes and blameworthy judgment-sensitive attitudes. He criticizes Scanlon's view of responsibility and suggests that responsibility may indeed bottom out in acts of choice rather than in our character. Moreover, Levy shows that volitionism can give a more plausible explanation of cases of character change and denies that whether the character change is temporary or permanent should be relevant to responsibility. Levy also argues that the inability of an agent to grasp a certain moral requirement may indeed be an excusing condition and that the 'in principle control' invoked by some attributionists is too weak to ground responsibility. Levy criticizes Angela Smith's account by disputing the idea that if it is appropriate to ask an agent to justify her action, this suggests that she is responsible for it. Finally, Levy points out how volitionism can more

accurately assess cases of apology.

Levy claims that when we say an agent controls some state of affairs, we are not simply saying that she can alter it, but also that she knows that she is causing those alterations and that she knows how she is causing them. “(...) the blade of grass we crush and bend as we walk, the air molecules we breathe in and out (...)”⁵⁵ are things with which we interact without controlling them, Levy notes. Thus, knowing what we are doing seems required for responsibility, if control is a requirement of responsibility. For the attributionist, the type of control sufficient for responsibility is the one the absence of which prevents us from attributing the attitude to the agent. Therefore, for the attributionist for the agent to be responsible for her action it is not required that she knows what she is doing because not knowing it does not make the action not attributable to her. Hence, the sense of control employed by the attributionist does not satisfy the epistemic condition for control that Levy claims to be crucial.

Both volitionists and attributionists agree that attitudes normally express the agent's commitments and judgments, and that some of these attitudes are bad whereas others are good. However, unlike attributionists, volitionists believe that not everything that is attributable to an agent is something for which she is responsible. Volitionists instead believe that attributing good and bad qualities to an agent is distinct from holding her responsible. Attributionists believe that if a bad quality is attributable to an agent, that agent will be blameworthy for it. Volitionists believe that attributing a bad quality to an agent is not equivalent to attributing blameworthiness, because there is a distinction relevant to responsibility which the attributionists do not find relevant.

The attributionist holds that the judgement-sensitive attitudes of the agent need not be controlled by the agent for them to be good or bad. And this is a claim the volitionist would agree with, because, for the volitionist, finding a moral flaw in an agent, for instance, has nothing to do with whether the agent can control it or not. There are negative assessments of morally significant features of the agent, which do not require control and do not entail blame. As the attributionist does not hold an agent responsible for a fault that results from a temporary mental illness, because they think it would not be attributable to the agent, similarly the volitionist does not hold an agent responsible for a fault over which she has no control (in a stronger sense – including epistemically). Levy thinks that attributionists should acknowledge this control condition on responsibility and that they should provide an argument to show that, with regard to judgment-sensitive attitudes, volitionists are mistaken in drawing a distinction between an attributable blameless judgment-sensitive attitude and an attributable blameworthy judgment-sensitive attitude.⁵⁶

Levy criticises Scanlon's account of responsibility. According to Scanlon, what a person is like can be seen from her attitudes. And if the agent's action expresses her attitudes, the state of the agent's character is visible through that action. For instance, consistently expressing laziness through my actions makes me lazy. Scanlon seems to believe that the character of the agent is the place our assessment of responsibility bottoms out, whereas volitionists such as Levy think that responsibility bottoms out with acts of choice. This means that for the attributionist what makes an agent responsible for an action is the fact that she is first responsible for a character trait, which is expressed in that action. In contrast, for the volitionist what makes an agent responsible for an action is the fact that she has chosen to

perform it. Levy is aware that Scanlon might object to this volitionist position by noting that since we are not responsible for our ability to choose, making the act of choice the place where responsibility bottoms out would lead to an infinite regress. However, Levy rightly points out that just because we are not responsible for our ability to choose, it does not follow that responsibility does not bottom out in the agent's act of choice.⁵⁷

Levy counters Scanlon's possible objection by noting that it is generally not up to us, for instance, whether we possess a certain normative competence, since it may depend on our upbringing and native capacity. But this does not entail that we can never be responsible for our actions. Actually, Levy suggests that there is a distinction that the attributionist cannot make but that can be made under the volitionist account, which is that we can *sometimes* be responsible for (aspects of) our character while at other times we are not. Imagine a situation where by taking a drug I cause a (temporary) change in my character and I make myself no longer normatively competent. In this case – Levy claims – it indeed matters to responsibility whether the agent satisfies control-conditions over her normative competence. Attributionists might only be able to respond that an agent is always responsible for her normative competence, just because it is a central part of her character, regardless of the control she has over that competence. Attributionists might believe that our character is always an accurate reflection of itself, that our practical identities are always accurately reflected by it, and that is why we are always responsible for it.⁵⁸

However, according to Scanlon, who is an attributionist, temporary changes in the agent's character, resulting from mental illness or brain damage, make the

57 *Ibid.*, p. 7.

58 *Ibid.*, pp. 7-8. The disagreement may also be prompted by an ambiguity in defining what character is.

attitudes produced not attributable to the agent. So it looks as if in cases of temporary change of character, the agent will not be responsible for his changed character, on Scanlon's view. However, Scanlon claims, if the character change is permanent and the agent does not reject the new attitudes, these can be attributed to her and she is then responsible for her new permanent character. Because of the brain damage or mental illness, the agent will be unable to control his attitudes and to see what responses are appropriate. But for Scanlon this fact is not relevant to the agent's responsibility. By contrast, Levy thinks this should be taken into consideration because, again, an attitude over which the agent has no control might be bad but not blameworthy.

Levy asks us to imagine that a character change were caused by an evil neuroscientist. The attributionist would insist that the agent is responsible for her actions which reflect her new attitudes stemming from her new character, if that change is permanent. But if we suppose that the neuroscientist has the means to reverse the character change, then "whether her victim is blameworthy at t depends upon whether the neuroscientist reverses the operation at t_1 " (which is later than t).⁵⁹ But this means – as Levy notes – that the agent's responsibility depends on factors "such (...) as the state of the neuroscientist's fridge, in which she has stored the brain tissue she has removed".⁶⁰ And it seems false that an agent's responsibility could depend on such factors.

Here Levy's objection might not be completely on point, because for attributionists what matters to responsibility may be not whether the character change is reversible but rather whether it is in fact reversed. However, if we consider two

59
60

Levy (2005), p. 8.
Ibidem.

scenarios, one in which the neuroscientist reverses the character change at t_1 and one in which he doesn't, it still seems counter-intuitive to claim that in the first case the agent was blameless at t , whereas in the second case she was blameworthy. The reason is that it does not depend on the agent whether that character change will be permanent or not. In general, it is not clear why the persistence of a character trait should matter to responsibility, as Scanlon argues. The fact that the agent does not reject her character change could play a role in explaining her responsibility only if she realises that she has changed in the first place, and that she now has a bad character trait. After all, if the agent does not realise that she has a new character trait and that the character is bad, she will not even be in a position to reject it, and she will not then be blameworthy for not rejecting it.

Levy then offers the example of a psychopath who cannot grasp that "(...) that an action will hurt someone always counts as a (defeasible) reason against it".⁶¹ Scanlon would say that this incapacity is not what would make the psychopath blameless for hurting someone and that the fact that the psychopath does not value harms is what makes us conclude that his character is flawed. However, Scanlon seems to overlook the distinction between simply contravening a requirement and disregarding it. A requirement can be disregarded only if the agent grasps it. The volitionist thinks that an agent is blameworthy only if she actively disregards a requirement, and not simply if she fails to meet it. Levy's claim is that not only must the agent know about the requirement, but she must also have reasons to think it should be respected. The psychopath cannot challenge the moral demand which he does not grasp. Thus, I agree with Levy that moral ignorance can excuse from responsibility:

61

Levy (2005), p. 9.

Suppose that there is a kind of harm that is objectively morally relevant, but of which we are ignorant (...) for instance, that plants can be harmed, and that this harm is a moral reason against killing or treading on them. In that case, many of us are causally responsible for a great many moral harms.⁶²

Levy claims that since we do not know about that moral demand and we are blamelessly ignorant, although we are causally responsible for killing or treading on plants, we are not blameworthy for it. This suggests that there must be a kind of epistemic condition that bears on responsibility which attributionists do not seem to accept. And they do not accept it because the absence of this condition does not undercut the attributability of our actions as following from our judgment-sensitive attitudes. The attributionist only allows for weaker conditions: it is enough that our judgment-sensitive attitudes are *in principle* controlled by the agent. And since we do not find ourselves with our judgment-sensitive attitudes as we do with our height, the attributionist thinks that we are somehow active with regard to our judgment-sensitive attitudes, as they are the result of the exercise of our agency. Therefore, the attributionist concludes, we can be properly blamed for these attitudes.⁶³

I agree with Levy that it is hard to see why the fact that judgment-sensitive attitudes are in principle under the control of reason could be a reason why we are responsible for them – as Scanlon claims. Levy rightly objects: “I don’t have a kind of ersatz control over my car if the steering wheel falls off; the fact that cars are in principle controllable does not alter my lack of control in that particular

62

Ibidem.

63

Levy (2005), pp. 9-10.

circumstance”⁶⁴. Indeed, it is not clear why a kind of control that, when absent, does not cancel responsibility would matter and why we should appeal to it.

Attributionists believe that a moral view held by an agent differs from her height because a moral view is, in principle, sensitive to judgments, whereas height is not. This ‘in principle control’ is not further spelled out, but Levy guesses that the attributionist claim could be countered as follows: “(...) suppose we discover that human beings are a rarity in the galaxy: one of the very few intelligent species whose height is not judgment-sensitive.”⁶⁵ If this were the case, according to attributionists, we would be responsible for our height, because the fact that humans cannot control their height would not be an excusing condition. But since it seems wrong to say that we would be responsible, it follows that whether we have in principle control is irrelevant to responsibility. And – as Levy points out – even if it were true that we can control our height, as long as we do not know how to control it, we would not be responsible for not doing so, since the control relevant to responsibility requires knowledge.⁶⁶

According to Angela Smith, the sense in which we are active with regard to our judgment-sensitive attitudes is that we can influence them when a justification for them is demanded.⁶⁷ If no reason is made available to support them, they will likely weaken or even cease to exist. Levy, in criticising Smith’s account, points out that the problem with claiming that moral criticism is based upon this type of activeness is that, when we call for justification, we are already drawing the attention

64

Ibid., p. 10.

65

Ibidem.

66

Levy (2005), pp. 10-11.

67

Smith (2005), p. 258.

of the agent to the attitude we are asking her to justify and which can be (in principle) corrected by the agent. But suppose that an agent causes an undesirable state of affairs without knowing that she has caused it or/and that it is an undesirable state of affairs. If it is in her power to change it and make it better, it seems appropriate to ask her to provide justification for her action and to modify that state of affairs, even without attributing blameworthiness for causing it. When we ask the agent to justify a bad attitude, we direct the agent's attention to that attitude, thus contributing to bringing about those epistemic conditions which volitionists demand to be satisfied in order for the agent to be morally responsible. And the reason why volitionists demand the satisfaction of these conditions is that they are essential for the kind of control relevant to responsibility. It is because of this that asking for justification for an attitude is also contributing to bringing it under the control of the agent. The control which makes her responsible for the attitude. But this does not imply that she was responsible all along before she was asked to justify or modify her attitude.⁶⁸

Smith claims that it is always justified to ask an agent to apologise for past bad judgment-sensitive attitudes, as a way to acknowledge her responsibility. However, as Levy notes, apologies can be used either to acknowledge responsibility or to express its absence.

If I step on your foot on a crowded subway, we shall both believe that an apology is called for. (...) "I'm sorry", I might say. "Just for a moment I thought it might be fun to step on someone's foot. Please forgive me". Much more commonly, I might say, "Sorry, I didn't see you there" (...).⁶⁹

68

Levy (2005), p. 11.

69

Ibid., p. 12.

While in the first case I am acknowledging responsibility, and you have the right not to accept my apologies, in the second case I simply explain that I did not choose to step on your foot, and that, therefore, my action was not malicious. This means that what I am saying in the second case is that I am not morally responsible for stepping on your foot, and that moral criticism would not be appropriate. When we apologize in this second sense, we usually mention excusing circumstances (*e.g.* I was distracted by *x*) to at least weaken our acknowledgement of responsibility. If I am not responsible for my lapse when I forget my friend's birthday, my apology would not be an acknowledgement of responsibility. As Levy suggests, if there was nothing that I did or refrained from doing to cause my lapse, I do not owe my friend an apology that expresses responsibility, but maybe only one that expresses my refusal to identify with a person who does not care about her friend's birthday.⁷⁰

As a final remark, in order to support the volitionist perspective further, Levy recalls what Watson claims about aretaic appraisals. Watson maintains that in order to be appraised in terms of a certain normative concept, an agent must possess that concept; for instance, an agent must know what cruelty is to be appraised as cruel. It is clear to Levy that that cannot be sufficient for the agent to be responsible for her cruel actions. In order to be responsible the agent must also be aware that the concept of cruelty applies to her action in that particular situation. Referring back to the case of plants suffering harm imagined by Levy, we are not responsible for cruelly plucking roses if we are not aware that they suffer harm. Not knowing that means not being able to control our action, and that is why adopting certain ends does not imply

taking responsibility for them, contrary to what the self-disclosure view holds. Levy claims that since we have no reason to take seriously the possibility that plants suffer harm, we do not take a stand on the issue of cruelty to plants. Taking a stand means asserting the superiority of a certain end over alternatives that I take seriously. Adopting an end does not then imply, by itself, that I take a stand for it, and therefore, that I am responsible for what I do accordingly.⁷¹ The self-disclosure view seems to reflect the attributionist position. Attributionists, such as Scanlon, would claim that our incapacity to see that plucking roses is cruel (in this scenario) is not what makes us blameless for being cruel in plucking roses. For if our action is an expression of us not valuing harms, then we might be blameworthy for cruelly plucking roses. And this seems false to a volitionist.

4.0

In conclusion, in this chapter I have explained some differences between attributionism and volitionism. In order to do so, in 1.0 I have drawn from Watson's distinction between the aretaic and the accountability faces of responsibility. Then, in 2.0 and 2.1, I have presented two examples of attributionist account defended, respectively, by T.M. Scanlon and Angela Smith. Finally, in 3.0, I have outlined Levy's considerations against attributionism and in favour of volitionism, and presented his objections to some of the claims made by Scanlon, Smith and Watson. It seems to me that Levy is successful in showing that volitionism is superior to attributionism. I have given sufficient reason in favour of volitionism and against attributionism to motivate further articulation of volitionism. I agree with Levy that

a stronger control condition is required for moral responsibility, and that not only the agent must have control over performing the action, but she also must know what she is doing in performing it.⁷² This condition allows us to draw the appropriate distinction between bad and blameworthy. It also seems to me that volitionism gives a better explanation, in terms of responsibility, of cases of character change, and it assesses more accurately cases of apology. The task of the following chapters will be to articulate the epistemic aspects of the control condition.

⁷² I disagree with Levy about how to characterize this epistemic condition. This will be discussed in chapter 6.

Chapter 2 – Culpability and Belief

In this chapter I will explore the question of whether and when it is appropriate to blame someone for holding a belief. Volitionism claims that blaming for holding a false belief is appropriate only if we have (direct or indirect) voluntary control over that belief.⁷³ A volitionist who believed that we lack such control over our beliefs would claim that blaming someone for holding a false belief is inappropriate. On the other hand, attributionism claims that holding a false belief may be an epistemic vice which reveals a defect of our character.⁷⁴ And this makes us an appropriate object of blame for holding that belief.

I will claim that it is appropriate to blame an agent for a false belief if she chose not to inform herself and she knew she should, to avoid forming a false belief. My view differs from that of those volitionists who think that beliefs can be blameworthy only if we have direct control over them, since according to my view this kind of control over the belief is not necessary for culpability. My position also differs from attributionism because I claim that what makes the agent an appropriate object of blame for holding a false belief is her decision not to inform herself, and not the viciousness of her character that that decision reflects.

In this chapter I will evaluate two positions: van Woudenberg's and McCormick's. While van Woudenberg's position seems to line up more with a volitionist view, McCormick's position is closer to attributionism. Van Woudenberg

73

See an example of such a view in Levy (2007).

74

See an example of such a view in Owens (2000).

assumes that force and ignorance, as commonly acknowledged, are excusing conditions for action. I will focus on ignorance. Van Woudenberg's thesis is that ignorance that excuses agents for performing wrongful acts can also excuse them for holding false beliefs. But not all false beliefs are excused by ignorance. And this gives reason to think that we can be blameworthy not only for our acts but also for our false beliefs. I will not deny this thesis. I will instead proceed as follows: in section 1.0 I will look at van Woudenberg's identification of three ways in which we fail to know a proposition and at his assessment of *whether* and *when* those are cases of excusing ignorance. Without disputing his classification, I will criticise his evaluation of cases of unwarranted or unjustified belief. In section 1.1 I will take into consideration van Woudenberg's conditions of culpable ignorance and I will criticise his analysis of cases of culpable and non-culpable ignorance. I will conclude by providing my alternative proposal to van Woudenberg's formulation of the conditions of culpable ignorance.

As far as McCormick's view is concerned, she argues that there is a kind of control we have over our beliefs that makes it appropriate to blame for holding them. This consists in owning our beliefs (and the mechanisms from which they result) and in taking responsibility for them. In 2.0 I will consider this position. To support her thesis, McCormick assesses cases of mistaken perception, memory, and practical reasoning. I will object to her assessment of those cases and conclude that the conditions she takes to be sufficient are only necessary for appropriately blaming an agent for holding a belief.

I have chosen to examine van Woudenberg's account because his

classifications of different ways in which we fail to know something is helpful to get an overview of the range of cases we must assess when we consider the relation between moral responsibility and ignorance. On the other hand, the reason I have chosen to discuss McCormick's view is that the cases she illustrates are very useful to effectively explain my opposition to views such as hers.

1.0

In this section I will look at van Woudenberg's identification of three ways of failing to know a proposition and at his assessment of whether and when those are cases of exculpating ignorance.

In his paper “Excusing Conditions for False Beliefs: Ignorance and Force”, van Woudenberg assumes that failing to know a proposition p has been widely recognised as a condition under which we *may* be excused for failing to perform an act we should have performed.⁷⁵ I may fail to know a true proposition p in different ways:

(1) *I unwarrantedly or unjustifiably believe that p .*⁷⁶

(2) *I do not believe that p .*

(3) *I believe that not- p .*

I will examine each of these three ways, but my criticism will focus on (1).

An example of (1) provided by van Woudenberg is the following:

75

Let us take acts to include both actions and omissions.

76

I will leave aside the distinction between 'unwarranted' and 'unjustified'.

[Suppose you] open a door, thereby killing a puppy that (...) was hiding behind it. (...) Now suppose you didn't know that the puppy was hiding behind that door because your true belief lacked warrant or justification – suppose, for example, that you have that belief because you misheard what someone was whispering to you. That person said, 'There is a pumpkin on the floor,' but you hear her saying, 'That puppy is behind the door'.⁷⁷

You believed there was a puppy behind the door and you did not refrain from opening it. Therefore – van Woudenberg claims – you are blameworthy for the puppy's death.

I agree with van Woudenberg that the necessary condition pertaining to belief requires mere belief, not justified belief. However, I find his analysis of this specific case inaccurate. Given how the example is presented and assessed by van Woudenberg, it seems that what makes you blameworthy for the puppy's death is just believing that the puppy is behind the door. However, it seems to me that van Woudenberg fails to take into account the possibility that you *may* still be blameless for the puppy's death if you did not know that opening the door would cause the puppy's death.⁷⁸ Now let us suppose that you did know that your opening the door would cause the puppy's death. Even in this case you may still be blameless for the puppy's death if you did not know that killing the puppy is impermissible.⁷⁹ For instance, it may be the case that your internal reasons⁸⁰ to walk through the door outweigh those not to kill the puppy. And it would not be fair to blame you in this

77 Van Woudenberg (2009), pp. 373-374.

78 See chapter 5 for more discussion on the foreseeability condition of moral responsibility.

79 See chapters 5 and 6 for more discussion on the thesis that an agent is blameworthy for an act only if he believes he should not perform it.

80 By internal reasons I mean only the reasons I do contemplate, the reasons I think I have for performing an action.

case, since what you did was simply to act according to your genuine better judgement, which is all we can be expected to do as agents with limited knowledge, when we deliberate: to do our (moral) best with what we know. This is one of my crucial claims about blameworthiness. I will come back to it in the next chapters.

Attributionists may respond that an agent is blameworthy for lacking the belief that killing a puppy is impermissible *if* this expresses a lack of care (attributable to the agent) for something a rational agent should care for. Indeed there are cases in which ‘not believing that killing puppies is impermissible’ would be blameworthy. But it seems to me that the lack of care would be relevant to the agent’s blameworthiness concerning her action only if the agent had developed lack of care for puppies’ life as a result of some earlier choice over which she had control. Suppose I believe that killing puppies is impermissible. Notwithstanding this, I am in a situation in which I have self-interested reasons to kill a puppy. I can do it without feeling guilty afterwards by taking a pill that will temporarily make me believe that killing puppies is permissible. If I take the pill, and kill a puppy, I am then indeed blameworthy for taking the pill and for holding the belief that killing puppies is permissible. And the reason is that I knew this would be the result of my taking the pill. Moreover, I am also blameworthy for killing the puppy because that was the aim of my taking the pill.⁸¹ In deciding to take the pill, in a way I express lack of care for the life of puppies. But the fact that indicates that I deserve blame is that I have taken the pill intending to prevent the belief that killing the puppy is impermissible from interfering with my action, so that I kill the puppy (for my self-interested reasons)

81
of responsibility.

See chapter 5 for more discussion on the ‘cognitive connection’ condition

without feeling guilty afterwards.

The second way in which – according to van Woudenberg – I can fail to know that *p* is the following:

(2) *I do not believe that p.*

However, there is a further distinction to take into account between (2a), (2b), and (2c):

(2a) *I neither believe nor disbelieve p, having entertained p.*

And (2a) can be divided again into (2aa) and (2ab):

(2aa) *I neither believe nor disbelieve p (having entertained p) when someone else knows that p.*

For instance, “S may not know the answer to the question, 'Who wrote *De Mare Librum* [sic]?' but somebody else does know it”.⁸² This ignorance can be culpable or non-culpable, and it can excuse only if it is non-culpable – van Woudenberg claims.

(2ab) *I neither believe nor disbelieve p (having entertained p) when nobody else knows that p.*

There is also a further distinction between (2aba) and (2abb):

(2aba) *I neither believe nor disbelieve p (having entertained p) when nobody else knows that p, but when somebody could easily get to know that p.*

S may not know the answer to the question 'Is there a puppy behind the door?' and as a matter of fact nobody knows it at that time (there is simply nobody around), but the answer

can be very easily known – a look in the right place will suffice.⁸³

Again, this ignorance can be culpable or non-culpable and, according to van Woudenberg, it can excuse only if it is non-culpable.

(2abb) *I neither believe nor disbelieve p (having entertained p) when nobody else knows that p, and nobody else could easily get to know that p.*

Van Woudenberg considers the following case:

(...) you have sworn a dire oath to the members of a club you belong to, to bring relief to the poor. A reliable agent has directed your attention to the poor Smith family. You give them a handsome amount of money, only to find out later that just as you have handed them the check, they are announced the winners of the local lottery. Your gift to the Smith family thus has a bad effect: you haven't really brought relief to a poor family, but instead made a family that was soon to be fortunate, even more fortunate.⁸⁴

It was impossible to know the winner of the lottery in advance. Therefore, you are blameless for your action – van Woudenberg claims. Being ignorant of something impossible to know will always be non-culpable. Therefore, (2abb) will always excuse me.

I agree with van Woudenberg that the kind of ignorance mentioned in the example always excuses. However, the fact that it was impossible for anyone to know the winner of the lottery in advance is not what makes you blameless for your ignorance. It is true that the impossibility of knowing the winner of the lottery in advance

83

Ibidem.

84

Ibidem.

guarantees that you will not be culpable for your ignorance. Still, it is the non-culpability of your ignorance that makes you blameless for giving the money to the Smith family, not the impossibility of knowing the winner of the lottery in advance. This suggests that it is not the fact that something is more or less difficult to know that makes the agent responsible for not knowing it. Rather, what makes the agent responsible is whether that failure to know is culpable or non-culpable.

(2b) *I neither believe nor disbelieve p, never having entertained p.*

S has never been interested in Dutch intellectual history, nor in legal history, and accordingly has never heard of Hugo Grotius. So he has no awareness of the question, 'Was Grotius the author of *De Mare Librum* [*sic*]?' Accordingly, he doesn't believe the proposition that Grotius is the author of *De Mare Librum* [*sic*]; but neither does he disbelieve it⁸⁵.

Going back to the case of the puppy, if you have never entertained the proposition that there is a puppy behind the door, you seem blameless for the puppy's death only if this ignorance is non-culpable – van Woudenberg asserts. Therefore, (2b) excuses if the ignorance is non-culpable.

However, this kind of ignorance can never be culpable. If I chose, for example, to take a pill that makes me never entertain the proposition that there is a puppy behind the door, in order to do so I would have to consider such a proposition in the first place. Thus, I would not be ignorant. Alternatively, I could take a pill that makes me never entertain any belief. But it seems too harsh, in this case, to claim that I would be culpable for every future belief I should have. It seems that only if I foresee that I might kill the puppy I

would be culpable for not holding the belief that the puppy is behind the door (this foreseeability condition will be discussed in chapter 5). And if I foresee that I will kill a puppy that is behind the door, this means that I have entertained the proposition that there is a puppy behind the door. So, again, I am not ignorant.

(2c) *I disbelieve p.*

Van Woudenberg notes that disbelieving *p* requires entertaining *p*.

(...) Sam cannot be said to disbelieve the proposition that storks bring babies when he has never so much as entertained that proposition, but he can be said to disbelieve the proposition that $2+2=5$ when he *has* considered that proposition – and rejected it⁸⁶.

Here van Woudenberg focuses on false propositions. If you disbelieve that storks bring babies you are not ignorant, since it is false that storks bring babies. Thus, van Woudenberg concludes, disbelieving false propositions is *not* ignorance.

Finally, the third main way of failing to know that *p* is the following:

(3) *I believe that not-p.*

Suppose you believe that Grotius is not the author of *De Mare Librum* [*sic*]. Then you lack knowledge because the proposition you believe is false. (...) [And you are] also ignorant with respect to the proposition that Grotius is the author of *De Mare Librum* [*sic*].⁸⁷

The belief that not-*p* can be culpable or non-culpable. The belief that not-*p* can excuse only if it is non-culpable. Furthermore, van Woudenberg maintains, the

86

Ibid., p. 375.

87

Ibidem.

belief that not-*p* excuses only if it is relevant to the action or its unwanted result. If I have the belief that Rome is not a city, this belief is not relevant to my killing the puppy that is behind door or to the puppy's death. It is instead relevant, for instance, whether I hold the belief that the puppy is behind the door.

The following table summarises van Woudenberg's classification and evaluations:

Ways of failing to know <i>p</i>:	Excusing/non excusing
(1) I unwarrantedly or unjustifiably believe that <i>p</i>	This ignorance does not excuse.
(2aa) I neither believe nor disbelieve <i>p</i> , having entertained <i>p</i> , when someone else knows that <i>p</i> .	This ignorance excuses only if it is non-culpable.
(2aba) I neither believe nor disbelieve <i>p</i> , having entertained <i>p</i> , when nobody else knows that <i>p</i> , but someone could easily get to know that <i>p</i> .	This ignorance excuses only if it is non-culpable.
(2abb) I neither believe nor disbelieve <i>p</i> , having entertained <i>p</i> , when nobody else could easily get to know that <i>p</i> .	This ignorance always excuses, since it is always non-culpable.
(2b) I neither believe nor disbelieve <i>p</i> , never having entertained <i>p</i> .	This ignorance excuses only if it is non-culpable
(2c) I disbelieve <i>p</i> .	This is not ignorance when <i>p</i> is false.
(3) I believe that not- <i>p</i> .	This ignorance excuses only if it is non-culpable and if the belief that not- <i>p</i> is relevant to the action or its unwanted result.

I did not dispute this classification. My most substantial remarks have been about van Woudenberg's evaluation of the example provided for (1). I have claimed that the main problem with van Woudenberg's view is that he assesses as

blameworthy cases that need to be described further in order to be fairly evaluated. Van Woudenberg overlooks distinctions that are relevant to whether the agent deserves blame. In particular, he does not consider whether the agent is aware of what their act would cause and of the impermissibility of the act itself. Consequently van Woudenberg's assessment encompasses as blameworthy cases which are instead blameless.

1.1

In this section I will discuss the question of when ignorance is culpable by considering van Woudenberg's account of culpable ignorance. In doing so I will criticise the analysis of two cases he examines. Here are van Woudenberg's conditions of culpable ignorance:

(CI) *S*'s ignorance with respect to *p* at time *t* is culpable when (a) given *S*'s position at *t*, *S* should have known that *p* at *t* and (b) given *S*'s abilities and circumstances at *t*, *S* could have known that *p* at *t*⁸⁸.

Occupying a certain 'position' entails knowing certain things. *S* fails to fulfil *S*'s duty if she does not know what the position requires one to know. Van Woudenberg notes that it may be the case that someone cannot know what she should know, because of abilities or circumstances.⁸⁹ In other words, van

88 *Ibid.*, p. 377.

89 A surgeon, because of the position she occupies, should know how to successfully perform heart surgery. But suppose she suffers from a brain injury which undermines her medical abilities or suppose that she is not told that the patient is allergic to the anesthetic used during

Woudenberg denies that “ought”, when expressing an obligation, implies “can”.

Furthermore, van Woudenberg mentions another condition for culpability valid only for *false beliefs*:

(CI') *S*'s ignorance, consisting in having the false belief that *p*, is culpable if (c) *S*'s belief that *p* is unjustified.⁹⁰

For van Woudenberg lack of justification, on whatever account of justification, makes a false belief culpable (the less justified a belief is, the more it is culpable to hold it).

Then van Woudenberg examines, according to his conditions of culpable ignorance, a case of *blameless* ignorance:

Suppose you left a check on your desk, then naturally and typically you believe that that check is on your desk. You will continue to believe so, of course, even when, unbeknownst to you, a thief broke into your house and stole everything that was on your desk⁹¹.

In this case, you have never entertained the proposition that someone stole everything from your desk. You neither believe nor disbelieve the proposition that someone stole everything from your desk (see (2b)). The truth of the proposition that someone stole everything from your desk makes the proposition you believe (*i.e.* that that

the surgery. Because of this she fails to perform heart surgery successfully. There is a sense in which she should have acted differently although she could not know how to do so. I believe this characterization is still ambiguous when discussing issues of moral responsibility. See chapters 3 and 5 for more discussion of the sense of 'should' relevant to moral responsibility.

90

Van Woudenberg (2009), p. 377.

91

Ibid., p. 379.

check is on your desk) false. But you neither should have nor could have known about the thief's action. You have no reason to make sure that your check is still on your desk. Your false belief is justified. Therefore, van Woudenberg concludes, you are blameless for your belief.

Then van Woudenberg moves to what he believes to be a case of culpable ignorance. I will explain why I find his analysis of the case puzzling.

Suppose Sam believes that his financial situation is healthy ($=p$). Suppose furthermore that his bank has sent him a letter telling him that he is bankrupt ($=q$) – a letter that Sam hasn't yet seen because it was delivered at the wrong address. (...) suppose that bank employees have communicated, prior to sending the fatal but undelivered letter to Sam, that his finances weren't at all going well ($=r$) and that Sam was alert, awake, and in good shape when they did.⁹²

Proposition r is true. Sam is aware of r . Proposition r *obviously*⁹³ entails that p is false. Van Woudenberg concludes that Sam is culpable for believing p .

As I said, I find this analysis puzzling. I do not see why the fact that proposition r *obviously* entails that p is false should be seen as relevant to Sam's culpability for believing p . The question to raise should be rather: 'Why does Sam believe that his financial situation is healthy although he is aware of r ?' For either he knows that r entails that p is false or he doesn't. And if he sees that entailment, then what stops him from believing that p is false? He might fail to believe that p is false because of a

92 *Ibid.*, pp. 379-380.

93 What van Woudenberg means here is that the harder it is for one to see the entailment, the less culpable one is for believing that p .

cognitive failure. And if he does not have any control over this cognitive failure, he is not culpable for believing *p*. If he does not see that *r* entails that *p* is false, then it seems to me that the mere fact that the entailment could be easily seen by anyone else does not make him culpable. Even if every rational being could see that entailment except Sam, that does not entail that he should be morally expected to see it, because the reason why he does not see that entailment may be out of his control. What is relevant to blameworthiness is whether Sam did not see the entailment *culpably*. And Sam would be blameworthy for not seeing the entailment only if there were a prior episode in which he chose not to develop the ability of seeing that entailment, knowing that this would excuse him from believing *p*. If there were no such episode – and I suspect this would be most often the case – Sam would be blameless.

Then van Woudenberg goes back to the case of the check, modifying it in order to turn it into what he believes to be a case of culpable ignorance. As with van Woudenberg's analysis of the previous case, I will criticise his analysis of the new version of the case.

(...) suppose that you knew that your wife would take the check right after you left the house (perhaps she thought it is far too dangerous to leave the check there, given the thefts that have been taking place in the neighborhood recently).⁹⁴

The proposition that your wife would take the check right after you left the house is true. The truth of this proposition implies that your belief that the check is on the desk is false. Therefore, van Woudenberg concludes, you are culpable for

believing that the check is on the desk.

However, it seems to me that the fact that the proposition that ‘your wife would take the check’ entails that your belief that ‘the check is on the desk’ is false is not relevant to your culpability for believing that the check is on the desk. The reason is that, again, to establish whether you are culpable we first need to ask: why do you believe that the check is on the desk although you know that your wife would take it? Either you know that the proposition that your wife would take the check right after you left the house entails that your belief that the check is on the desk is false or you don't. If you see that entailment, what stops you from believing that the proposition that the check is on the desk is false? If what stops you is some cognitive failure over which you have no control, you would be blameless for holding the belief that the check is on the desk. But even if you do not see that entailment, it seems to me that you are still not culpable for believing that the check is on the desk, regardless of the fact that the entailment could be easily seen by anyone else. Because to fairly attribute culpability to you we need to consider what *you* do not see and whether *you* do not see it culpably or non-culpably. The fact that you don't see the entailment but you *could* see it, considering your cognitive abilities, does not entail that you are blameworthy. It only suggests that since you are an agent with certain cognitive abilities, it is more likely that you will notice certain entailments. According to my volitionist account, you would be blameworthy for not seeing the entailment only if there were a prior episode in which you chose not to develop the ability of seeing that entailment, knowing that you should not do it. In all the other cases you would be blameless.

Then van Woudenberg further modifies the case of the check so that it

becomes, in his view, a case of blameless ignorance. I will reiterate my criticism of van Woudenberg's analysis for this version of the case: "Suppose that you believe that the check is on your desk ($=p$) but you *do* consider the proposition that someone stole everything from your desk ($=q$)".⁹⁵ Someone stole everything from your desk. You have the false belief that the check is on your desk. You are ignorant of the proposition that someone stole everything from your desk. If your ignorance is non-culpable and there is no true proposition r that you believe and that would entail the denial of your belief that p (or there is such a proposition r but you are not culpable for not believing r), then – van Woudenberg claims – your false belief is excused by your ignorance.

Once again, it seems to me that whether there is a true proposition r that would entail the denial of your belief that p has little relevance to culpability. For even if there were such a proposition and you believed it, it might be that you do not see that r entails the falsity of p , and if so, you are *not* culpable for holding your false belief, unless you are culpable for not seeing that entailment. But if you saw that entailment, then what would stop you from dropping your false belief? I am inclined to believe that, in most cases, it would be something for which the agent is not culpable.

To summarise my criticism of van Woudenberg: the main problem with his view is that he treats some cases of failure of rationality (*e.g.* Sam failing to see that r obviously entails the falsity of p) as cases that make the agent culpable for the belief she holds. The question that needs to be answered first is rather whether that failure was under the control of the agent, that is whether she brought it about as a result of some earlier choice she made, knowing that she should not do that because

it would (or might) lead her to hold a false belief. If van Woudenberg does not answer the question of whether the rational failure is culpable first, he will not be able to distinguish correctly between the cases in which the agent deserves blame and those in which the agent is blamelessly irrational.

Finally, van Woudenberg considers a version of the case of the check where the belief is unjustified.

(...) now suppose you believe that you have mailed the check – a check that, as you are well aware of, should be in the mail on pain of severe financial consequences for both your family and the company you work with – while in fact you left it on your desk⁹⁶.

Your belief is false and, according to van Woudenberg, if you have no justification for your belief, your having a false belief is not excused (on whatever account of justification). The more justification one has for a false belief, the more one's ignorance is excused by it.

However, it seems to me that van Woudenberg's assessment works only with a certain account of justification. In order to be blameworthy for believing that you have mailed the check, you must lack a particular kind of justification. You are blameworthy if you thought you had to look for more justification for your belief but you didn't, knowing that this could have led you to hold the false belief that you had mailed the check. What excuses is believing that you shouldn't look for more justification even though you believed that your belief that you have mailed the check could turn out to be false. What excuses is then believing yourself to have sufficient

justification in this sense.

I will now conclude by commenting on the conditions for culpable ignorance.

First, let us remind ourselves of van Woudenberg's formulation:

S's ignorance with respect to *p* at time *t* is culpable when (a) given *S*'s position at *t*, *S* should have known that *p* at *t* and (b) given *S*'s abilities and circumstances at *t*, *S* could have known that *p* at *t* (...). [In addition,] *S*'s ignorance, consisting in having the false belief that *p*, is culpable if (c) *S*'s belief that *p* is unjustified.

Van Woudenberg does not specify whether these conditions are necessary or/and sufficient. If they are only necessary, I agree with all of them. But if they are intended to be sufficient, as van Woudenberg's analysis of different cases suggests, then I disagree with van Woudenberg's account of culpable ignorance. From my objections to van Woudenberg's account, it can be expected that my alternative proposal would instead be as follows:

S's ignorance with respect to *p* at time *t* is culpable only when *S* should have known that *p* at *t*, but *S* previously chose not to inform herself about *p* knowing that she should have. Moreover, *S*'s false belief that *p* is culpable if *S*'s belief that *p* is unjustified, in the sense that *S* knows that *p* may turn out to be false and that she should seek more evidence to support *p*.

If van Woudenberg's conditions of culpable ignorance are intended to be sufficient then it seems to me that he is mistaken in not considering whether the agent chose not to inform herself knowing that she should do so to avoid being

ignorant (or holding a false belief). This is because a person may be in a situation in which she should have performed a certain action and she is, at the same time, blameless for not having performed that action.

2.0

In this section I will consider McCormick's position on whether beliefs can be culpable. First, I will look at the concept of control she takes to be relevant to blameworthiness for holding beliefs. I will explain what she means by taking responsibility for a belief and for the mechanism from which the belief results. Then I will look at cases McCormick presents where our perception, memory and practical reason are mistaken and I will criticise McCormick's assessment of these cases in terms of blameworthiness of the beliefs formed and maintained. My criticism focuses on McCormick's failure to show that the ownership of the mechanism from which our belief results is a sufficient condition of responsibility for that belief. I will suggest that her condition of responsibility may be, at most, a necessary condition, since if I do not doubt the efficiency of my perception, memory or practical reason, I cannot be appropriately blamed for not avoiding the resulting false beliefs.

Miriam McCormick, in her paper "Taking Control of Belief", claims that our practices assume that it is appropriate to attribute responsibility for holding beliefs. She brings as example exclamations such as "How can you believe that?" and "What a ridiculous thing to believe!" stating that these show that we hold the person in question responsible for holding that belief.⁹⁷ In other words, she claims that we see

97

I tend to disagree with this assumption. I would rather maintain that such

that person's belief as a consequence of her agency. McCormick acknowledges that we do not decide to hold a belief in the same way as we decide to perform an action⁹⁸, since there seems to be some lack of control in the case of beliefs which may lead us to think that we should be excused from responsibility. However, McCormick believes that there is a kind of control we possess even in forming and maintaining beliefs, called *guidance control*:

(...) imagine that you are driving a car that is a 'driver instruction' automobile with dual controls. As long as you are driving in a relatively safe manner, the instructor lets you control the car and so when you, at the correct time, turn to the right it is *you* who is guiding the car to the right. But if you had shown signs of confusion and were about to mistakenly turn to the left, the instructor would have stepped in and steered it to the right. Thus you could have gone in no other direction but to the right.⁹⁹

The example illustrates that you can have some kind of control over the car even if there is a 'fail-safe' mechanism thanks to which if you had failed to act, the car would have done the same as it did under your guidance. McCormick explains that to have guidance control over an action means that the mechanism which issues in the action is the agent's *and* is reasons-responsive. According to McCormick's conception, the mechanism is one's own when the agent takes responsibility for it, and this happens when one comes to view oneself as an agent. It happens when one realizes that one's decisions, intentions and actions have some effect on the world.

admonishments reveal merely that we disapprove what is believed and that we are upset with the fact that that person holds that belief. However, here I will proceed as if McCormick's assumption is true.

98

See chapter 5 for more discussion on the idea of choosing to believe.

99

McCormick (2011), p. 173.

For McCormick this goes together with the realization that one is a fair target of praise and blame. Thus, the agent takes responsibility both for the mechanism and for its resulting conduct. As the agent takes responsibility for the mechanism, he also takes responsibility for its present and future operations. McCormick also claims that it is when others blame or praise us that we develop a view according to which we take ourselves to be responsible agents. Taking responsibility, in this sense, has then to do with how we react to others and feel about ourselves. Taking responsibility, according to McCormick, is not a conscious act.

One problem I have with McCormick's view is that it does not allow for the following possibility: an agent may view herself as having an effect on the world with her decisions and actions, but she may also feel (and think) that she is not always a fair target for blame and praise. And that would be the case if, for instance, she found out that she was blamelessly ignorant about a relevant factor that affected her decision and action. Hence, it seems that McCormick fails to distinguish properly between causal responsibility and moral responsibility.¹⁰⁰ Moreover, I agree that experiencing blame and praise may lead us to view ourselves as responsible agents, but this does not necessarily mean that it is appropriate to *feel* responsible in all situations in which we are blamed and praised, for blame and praise may be unfairly attributed.

McCormick believes that these ideas of 'guidance control' and 'taking responsibility' can be applied not only to actions but also to beliefs. The mechanisms that result in beliefs are perception, memory, reasoning, attending, evidence

100
responsibility.

See Sartorio (2007) for a distinction between these two notions of

gathering, and inquiry. McCormick considers the following case:

(...) say I believe, on inadequate evidence, that my brother stole some money from me. I discover my error (perhaps by finding the money or finding the real thief). I will then (under normal circumstances) revise my belief about my brother¹⁰¹.

McCormick claims that in this case I can still feel guilty for holding that belief about my brother even though I never acted on it. And my feeling guilty reveals that I find myself responsible for the mechanism that results in my belief and for holding that belief. This shows then, according to McCormick, that we can be responsible for holding a belief.

However, it seems to me that the case is mistakenly assessed, since I do not see why I should feel guilty for forming the belief that my brother stole my money. I might indeed appropriately feel guilty for *maintaining* that belief if I think I should look for more evidence and I know that, by looking for more evidence, I might change my belief. But I shouldn't feel guilty for temporarily believing that my brother stole my money, for the evidence I think I have (though in fact inadequate) may suggest that it is my brother who stole the money. In general, since I have no direct control over forming a belief, but only over deciding to collect evidence that might make me change my belief, if I don't collect any more evidence – knowing that I should – I will not give myself a chance to change my belief. And that is what makes me blameworthy for then holding the false belief that my brother stole the money. This means that the mere fact that I formed the belief on inadequate evidence is not sufficient to make me blameworthy for holding the

belief, contrary to what McCormick seems to suggest.¹⁰²

McCormick continues with the following scenario: “Say I discovered that I had undergone some kind of psychic manipulation such that the mechanism responsible for issuing in beliefs about my brother was one which was controlled (via some remote control) by some evil scientist”¹⁰³.

McCormick believes that in this case I may still feel guilty for holding my belief, and that at least part of the reason is that the belief is the result of *my* agency. And this is because I have taken responsibility for the mechanism issuing in that belief. In this case I lost some control over that mechanism. The mechanism is now faulty and this is *my* fault. Here McCormick is not taking a clear stand on whether blame is appropriate in the modified case. She just wants to explain what matters in her account of culpability for beliefs. It seems to me that feeling guilty would be inappropriate in such a case, as the fact that my belief results from my agency and the fact that losing control over the mechanism issuing in the belief is a fault that belongs to me constitute only a necessary condition (not sufficient) for being culpable for holding the belief, as I shall show in what follows.

To strengthen her view, McCormick takes into consideration and evaluates two cases of mistaken perception, a case of mistaken memory, and a case of mistaken practical reasoning. I will criticise McCormick's assessment of the cases. The first example of mistaken perception is the following:

102 This applies also to a version of the case where I have inadequate evidence because I have no evidence. I am not directly in control of forming a belief. Hence, I can only be expected to look for evidence if I think that the belief I am holding is irrational. And if I decide not to look for evidence knowing I might maintain my irrational belief, then I may be blameworthy for maintaining that belief.

103 McCormick (2011), p. 175.

Imagine that someone is insistent that he sees a unicorn galloping toward him and based on the usual trustworthiness of his sense and [sic] believes there is a unicorn galloping toward him. (...) this person later discovers that a powerful hallucinogenic drug had been slipped in his drink and he still believes that a unicorn had galloped toward him (...).¹⁰⁴

McCormick claims that it is appropriate to criticise the person in the example because usually one is expected to revise one's belief if it is pointed out that one's mechanism has gone wrong. The criticism is appropriate as a reaction to one's being overconfident or hasty in the mechanism's functioning.

However, it seems to me that, once again, we cannot confidently assess the agent in the example as blameworthy if we don't first answer the question of *why* he still believes that a unicorn had galloped toward him after finding out about the drug. The example is then under-described. McCormick seems to suggest that the reason could be that he was overconfident or hasty in holding his belief. I take being overconfident about a belief to mean that I am certain that my belief is true even if there is in fact no (good) evidence to justify it. Now, either this is the case because I am not persuaded by the new evidence or because I have some cognitive defect that prevents me from moving to a new belief (*e.g.* that there was no unicorn). But in both cases, I do not see why I would deserve to be criticised, unless I deliberately brought about my not being persuaded by the new evidence or my being unable to move to a new belief.

To reinforce her case, McCormick provides a second example of mistaken perception:

Sam believes his girlfriend was kissing Jack in the car and he believes this because he saw her doing so. It turns out she was, in fact, not kissing Jack and somehow his perception was faulty and misled him into forming a faulty belief. (...) through pleading and insistence, Sam's girlfriend convinces him that she did not kiss Jack (...).¹⁰⁵

Again, McCormick claims that it is appropriate for Sam both to be blamed and to feel guilty for forming and maintaining the belief that his girlfriend was kissing Jack. The reason for this is that “there is some presumption that we should be ready to do some monitoring to ensure that this mechanism for which we have taken responsibility is operating correctly”.¹⁰⁶

However, my objection here is that ensuring that our mechanism is operating correctly is an action taken as a response to our doubting that our mechanism is not operating correctly. We cannot be expected to check it all the time. We can be expected to check it only if we have reason to think that it is flawed or that there is a lot at stake. For instance, we tend to give it more thought before accusing our partner of cheating, than before accusing them to have left the milk out of the fridge. This is because in the case of cheating there is more at stake. The question relevant to blameworthiness in the case of Sam is then the following: did Sam think he should doubt his perception considering his previous experience and what he believed was at stake? If the answer is yes, he may be culpable for not investigating the matter further and for maintaining his belief. But if the answer is no, he seems blameless for not doubting his own perception and for holding the false belief that his girlfriend was kissing Jack. Thus, according to McCormick, Sam may be culpable for not doubting his own perception even if he

105

Ibidem.

106

Ibidem.

himself did not believe that he should doubt it. In contrast, I claim that for Sam to be culpable for not doubting his own perception, it must be the case that he believed that he should doubt it.

Then McCormick moves to discussing a case of mistaken memory:

Imagine, you are conversing with a friend and you start talking about a conference you were both at the previous summer. You begin recalling together who else was at the conference. You say, 'oh and John was there – I remember liking his talk'. Your friend insists that John was not there and you are emphatic that you *remember* him being there. (...) the next day, your friend shows you the program and convinces you that you misremembered (...).¹⁰⁷

McCormick once again believes some blame is appropriate for holding the belief that John was at the conference. I disagree with McCormick that blame is appropriate here. After all, what I was claiming was simply that I *remembered* John being at the conference. Of course I could remember mistakenly, and you, my friend, simply claimed the opposite having, at the time, no more (internal) reason than me to support your belief. It is true that John was not there, but your reason to disagree with me did not seem stronger than mine, so I don't see why I should be criticised for not believing that you are the one holding the true belief. I may deserve blame if I thought I should doubt my memory more than yours and yet I insisted that I was right (just for the sake of winning the argument), but in this respect the example is under-described.

Finally McCormick considers a case of mistaken practical reasoning:

107

Ibid., p. 176. The emphasis is in the original.

(...) [Suppose] I am a juror and, through a process of deliberation come to believe the witness and acquit (...) [...] I am later convinced that there were good reasons to doubt the witness.¹⁰⁸

McCormick states that criticism – in the form of blame – is, once again, appropriate. My reason to object to this conclusion is that, as the example is presented, criticism seems most likely appropriate only in the form of the recognition of a flaw in the juror's reasoning, not in the form of blame. If, as a juror, I am not aware of such a flaw and I did not deliberately bring about my inability to see those good reasons to doubt the witness, I do not deserve criticism in the form of blame. We may want to say that it is unfortunate that I did not recognize the reasons to doubt the witness, but the example says nothing about whether this failure is culpable, and it is not appropriate to attribute blame just because it is *my* failure. It is a blameless failure if I came to believe the witness by following my better judgment and ignoring (blamelessly) the fact that there were better reasons not to believe them.

In conclusion, the main problem with McCormick's view is that although the ownership of the mechanism from which our beliefs result can be seen as a necessary condition of responsibility for that belief, McCormick fails to show that this is also a sufficient condition. My flawed perception, memory or practical reason are surely *mine*, but if I do not think I have most reason to doubt their efficiency, I cannot be appropriately blamed for not avoiding the resulting false beliefs.

3.0

In conclusion, in this chapter I have discussed van Woudenberg's and McCormick's views on the topic of culpability for beliefs. I have looked at van Woudenberg's identification of three ways in which we fail to know a true proposition: (1) unwarrantedly or unjustifiably believing that p ; (2) neither believing nor disbelieving p ; (3) disbelieving p . I have not disputed van Woudenberg's classification. However I have criticised his evaluation of a case of unwarranted or unjustified belief, by showing that van Woudenberg's conditions of culpable ignorance in such a case are not sufficient. I have then taken into consideration van Woudenberg's formulation of the conditions of culpable ignorance (CI+CI'). I have criticised his analysis of cases of culpable and non-culpable ignorance, by arguing that knowing a proposition r that entails the falsity of p is irrelevant to our culpability for holding p . Finally I have concluded by providing an alternative proposal to van Woudenberg's formulation of the conditions of culpable ignorance, interpreted as sufficient conditions: S 's ignorance with respect to p at time t is culpable only when S should have known that p at t , but S previously chose not to inform herself about p knowing that she should. Moreover, S 's false belief that p is culpable if S 's belief that p is unjustified, in the sense that S knows that p may turn out to be false and that she should seek more evidence to support p .

I have then taken into consideration McCormick's view on culpability for belief and I have criticised her evaluations of cases of mistaken perception, memory, and practical reasoning. I have conceded that the ownership of the mechanism from which our beliefs result is a necessary condition of responsibility for that belief, but

McCormick fails to show that it is also a sufficient condition. If I do not think I have most reason to doubt the efficiency of my perception, memory or practical reasoning, I cannot be appropriately blamed for not avoiding the false beliefs resulting from misperceiving, misremembering or misjudging.

Chapter 3 – Blameworthiness and Wrongness

In this chapter I will discuss the relation between the blameworthiness of the agent and the moral status of the action for which she is blameworthy. There are two main views in the literature on this matter. On the one hand, the *objective view*¹⁰⁹ states that an agent is blameworthy for performing an action only if the action is objectively wrong. On the other hand, there are philosophers who reject the *objective view* and claim that it is not necessary that an action be objectively wrong for the agent to be blameworthy for it.¹¹⁰ Ishtiyaque Haji and Andrew C. Khoury are among those philosophers who think that there are considerations that tell against the *objective view* and that blameworthiness does not require the objective wrongness, and it is affiliated only with the agent's mental states at the time of action.

In this chapter I will defend the *objective view* from some objections raised by Haji and Khoury. This is the main contribution of this chapter. I will not argue for the *objective view*. I will only suggest that Haji and Khoury do not successfully undermine the *objective view*. In addition, I will advance a view that combines the objective wrongness of the action with *certain* mental states of the agent, as follows: an agent is blameworthy for performing an action only if (a) she believes the action is wrong *and* (b) the action is wrong in a non-perspectival sense.

In 1.0 I will take into account two versions of a case presented by Haji to suggest that wrongness is not a necessary condition of blameworthiness. I will argue

109

Holly Smith (1983), (1991).

110

Brandt (1958).

that, in assessing the case, Haji misplaces the agent's blameworthiness. In 1.1 I will consider a Frankfurt-type case utilised by Haji to suggest blameworthiness does not require wrongness. I will criticise Haji's assessment of the case by disputing Haji's claim that the action in the Frankfurt-type case is *not* wrong. In 1.2 I will take into consideration two opposite cases examined by Haji to suggest that there isn't any affiliation either between blameworthiness and wrongness or between praiseworthiness and rightness. I will criticise Haji's assessment and claim that Haji is mistaken in suggesting that an agent can be blameworthy for performing a right action and praiseworthy for performing a wrong action. At the end of the section I will dispute Haji's positive proposal of a necessary epistemic condition of moral responsibility.

In 2.0 I will take into account three jointly inconsistent propositions on the relation between the blameworthiness of an action, its wrongness and the agent's mental states. I will discuss Khoury's attempt to reject one of the three propositions (in order to resolve the inconsistency), namely the one stating that wrongness is required for blameworthiness. I will then consider an example presented by Khoury to support his case and I will show that it fails to be persuasive in suggesting that we should reject the proposition that wrongness is required for blameworthiness. Moreover, I will dispute Khoury's claim that if we accept the proposition that wrongness is required for blameworthiness, we must give up the distinction between act-evaluations and agent-evaluations. Finally, in 2.1, I will take into consideration Khoury's epistemic condition of blameworthiness: an agent is blameworthy for an action only if the agent ought to believe that the action is wrong. I will consider some

examples presented by Khoury and assessed as cases in which the agent is blameworthy. I will show that in those cases there is at least one sense in which it is not the case that the agent ought to believe that what she is doing is wrong. I will also discuss Khoury's interpretation of "ought to believe" in terms of reasonable moral expectations on the agent, and I will suggest that, again, there is at least one sense in which, in the cases presented by Khoury, it might be unreasonable to expect the agent to believe that what she is doing is wrong. I will conclude by rejecting Khoury's epistemic condition of blameworthiness and by reaffirming my proposal as compatible with the *objective view*.

1.0

In this section I will take into account two versions of a case presented by Haji to show that blameworthiness does not require wrongness. In the first version of the case the agent performs a wrong action and believes that what he is doing is morally wrong. In the second version the agent performs a permissible action and believes that what he is doing is wrong. I will criticise Haji's assessment of the case regarding whether and for what the agent is blameworthy.

In his paper "An Epistemic Dimension of Blameworthiness", Ishtiyaque Haji presents the following case, which he calls *Mud Pie*:¹¹¹

(...) growls from his belly prompt Fred to pop his head into the refrigerator. His mouth waters at the sight of the pie on the top shelf; (...) Fred is perfectly aware that, given his

medical condition and certain other subsidiary factors, all things considered, he ought not to eat the pie. (...) Fred carefully embellishes the pie with dollops of cream and then proceeds (intentionally) to feast on it. (...) Fred could on this occasion have exercised self-control and refrained from indulging.¹¹²

This is a case in which Fred believes that his eating the pie is wrong. In addition, his eating the pie is in fact wrong. Haji claims that Fred is blameworthy for eating the pie.¹¹³ Haji maintains that an important part of what leads us to assess Fred as blameworthy is that he acted against his better judgement that he should not eat the pie. Then Haji offers an alternative version of the case, called *Phoney Pie*:

(...) suppose that Julia has concocted a pie that looks just like, and tastes just like, ordinary pie but with none of the properties exemplified by ordinary pie harmful to people suffering from the medical ailment that afflicts pie-loving Fred. (...) unbeknown to Fred, Julia has switched the ordinary pie with the benign look-alike, and it is now (morally) permissible for Fred to indulge. (...) Fred looks into the refrigerator, and acting against his all things considered judgement, gobbles the pie. (...) Fred (...) still acts contrary to his action-committing judgement that it is wrong for him to eat the pie, and he still exhibits the sort of control over his pie-eating that he exhibited in Mud Pie.¹¹⁴

This is a case in which Fred believes that his eating the pie is wrong. However, this time his eating the pie is permissible. Haji claims that Fred is still blameworthy for eating the pie. Haji states that *Phoney Pie* suggests that there is no

112 Haji (1997), pp. 524-525.

113 Haji is not concerned with blameworthiness for the consequences of one's action.

114 Haji (1997), pp. 525-526.

affiliation between the blameworthiness of an action and its objective wrongness. Instead, blameworthiness is associated with the fact that the agent acts taking himself to be performing a wrong action.¹¹⁵

I disagree with Haji's assessment of *Phoney Pie* and with his claim that there is no affiliation between the blameworthiness of an action and its objective wrongness. I understand Haji's idea of objective wrongness as wrongness in a non-perspectival sense. By this I mean that the wrongness of the action depends on the circumstances of the agent (*e.g.* his medical conditions) but it is independent of the agent's belief about the moral status of the action. The wrong action in a non-perspectival sense is then the action that Fred *should* not perform, in his circumstances.

It seems to me that in *Phoney Pie* Fred is not blameworthy for eating the pie, exactly because his eating the phoney pie is not wrong. Both the action "Fred's eating the phoney pie" and the action "Fred's not eating the phoney pie" are permissible (we stipulate). "Fred's eating the phoney pie" is indeed no worse than "Fred's not eating the phoney pie". From a consequentialist perspective, for instance, it is not the case that "Fred's eating the phoney pie" leads to worse consequences than "Fred's not eating the phoney pie".

Suppose I have to decide whether to paint my bedroom wall in yellow or green. Suppose I think that it is morally wrong to paint my bedroom wall in yellow but I do it anyway because I like the colour very much. It seems inappropriate to say that I deserve blame for painting my bedroom wall in yellow, because that is a morally neutral action. And its alternative was also morally neutral. The same applies

to Fred's case, in which both eating and not eating the phoney pie are morally neutral actions. However, here I am not claiming that Fred does not deserve blame for anything at all in *Phoney Pie*. My claim is only that he does *not* deserve blame for his eating the phoney pie, since there is nothing (morally) wrong with *that particular action*. Still, we might feel he deserves some blame, and I agree that he does. But what he deserves blame for is *the fact that* he acted against his better judgement that his eating the phoney pie is wrong, not for eating it *per se*.

I believe that philosophers who are interested in blameworthiness should be accurate not only in assessing *whether* an agent is blameworthy in a particular case, but also in establishing *for what* the agent is blameworthy. Acting against one's better judgement can be seen as a general action under which a great number of more specific actions fall. If wrongness *in a non-perspectival sense* is what we are concerned with, we should acknowledge that this kind of wrongness does not apply to the general action of acting against one's better judgment as well as it does not apply to other general actions (*e.g.* killing). Indeed we cannot say that it is wrong to act against our better judgement *per se*, because there will be cases in which our better judgement is mistaken and acting against it would lead us to perform an action that is permissible or even right. Hence, wrongness in a non-perspectival sense applies only to more specific actions. However, as Haji suggests, acting against one's better judgement happens to be also a blameworthy-making feature of particular actions. So when we blame Fred in *Phoney Pie*, we are not blaming him for his eating the pie, since that specific action is permissible. What we are doing is instead blaming him for the blameworthy-making feature of that action, *i.e.* the fact that he

acted against his better judgement.

It then seems to me that Haji is not successful in his attempt to show that blameworthiness does not require wrongness. He only appeals to the intuition that in *Phoney Pie* Fred is blameworthy for eating the pie. But this intuition seems misguided and Haji seems mistaken also on the content of the intuition itself. What Haji describes as the intuition that Fred is blameworthy for eating the pie is actually the intuition that Fred is blameworthy for the fact that he eats the pie against his better judgement. Thus, looking carefully at Fred's action, we realise that some attribution of blameworthiness is appropriate, but not for performing the action of eating the phoney pie, because that action is permissible. It is rather the fact that Fred acted against his better judgement which is the true object of the attribution of blameworthiness. Haji is then mistaken as he misplaces the attribution of blameworthiness.

In this section I have considered two versions of a case presented by Haji. In the first version, the agent performs a wrong action and believes that what he is doing is morally wrong. In the second case the agent performs a permissible action and believes that what he is doing is wrong. Haji judges both versions of the case as situations in which the agent is blameworthy for his action. And this would show that blameworthiness does not require wrongness. I have criticised Haji's assessment of the case by pointing out that Haji misplaces the agent's blameworthiness. In the version of the case where the agent performs a wrong action, he is blameworthy for it. In the version of the case where the agent performs a permissible action, he is blameless for performing that action although we may find him blameworthy for *the*

fact that he acted against his better judgement that he should not perform that action. This undermines Haji's objection to the objective view as he fails to show that blameworthiness does not require wrongness.

1.1

In this section I will consider a Frankfurt-type case discussed by Haji to suggest, again, that blameworthiness does not require wrongness. In the Frankfurt-type case the agent believes that what she is doing is wrong. Moreover, the agent could *not* have acted otherwise. I will criticise Haji's assessment of the case regarding the moral status of the action.

In his paper Haji examines the following Frankfurt-type case:

Imagine that, despite harboring the belief that squeezing the trigger (for the purpose of killing Max) is wrong, Jones – on her own – arrives at the decision to kill Max, and when the moment is right successfully (and nondeviantly) acts on her vile decision. Unbeknownst to her, had she shown any inclination not to intend or decide to murder Max, Dark Demon who had scrupulously been monitoring the situation all along, would have, *via* direct manipulation of Jones' brain, caused her to form an intention to squeeze the trigger and to act on that intention.¹¹⁶

Jones' action is under the control of her practical reasoning, it issues from her beliefs and desires, and it is performed in spite of Jones' belief that it is wrong. Because of this Haji claims that Jones deserves blame for squeezing the trigger. In

addition, Haji states that Jones' action is not objectively wrong, since the following principles are true:

(OW) It is morally wrong for S to perform [not to perform] A if and only if S has a moral obligation not to perform [to perform] A.¹¹⁷

(K) S has a moral obligation to perform [not to perform] A only if it is within S's power to perform [not to perform] A.

Haji also notes that the two principles above entail a third one:

(WAP) It is morally wrong for S to perform [not to perform] A only if it is within S's power not to perform [to perform] A.¹¹⁸

Haji concludes that since Jones cannot intentionally refrain from performing her action (*i.e.* squeezing the trigger), the action must be *not* wrong. Therefore, Haji believes himself to have shown again that it is not required for an action to be wrong in order for the agent to be blameworthy for it.

I disagree with Haji's conclusion because it seems to me that in Jones' case there is at least one sense in which the action that Jones performs can be seen as wrong in a non-perspectival sense. This is the sense of 'should not' that expresses moral undesirability. If we take this sense of 'should not', an action may be wrong (morally

117 Note that (OW) is controversial if we wanted to allow the possibility of cases where every option for an agent is wrong.

118 Haji (1997), p. 526. The principles have been reformulated to make the reasoning easier to follow.

undesirable) even if the agent that performed it could not act otherwise.¹¹⁹ We can see this sense of wrongness by noting that in the Frankfurt-type case we don't want the action of squeezing the trigger to happen (*e.g.* because it will bring more pain than happiness) if we compare it to its opposite (*i.e.* not squeezing the trigger). And even though Jones can't intentionally refrain from performing the action we may still think that that action should not take place. In other words, in those circumstances, that action should not be performed by Jones, or by any other agent in Jones' place. Actually, we don't want either that Jones perform the action intentionally or that Dark Demon makes Jones perform the action. And in this sense the action is then objectively wrong, because it is morally undesirable. From a moral point of view, we believe that agents *should not* perform morally undesirable actions. And, after all, morally wrong actions are exactly those actions we should not perform in morally relevant situations (*e.g.* situations where we might harm someone). Consequently, it seems that in order for an agent to be blameworthy, the action must be at least morally undesirable, which is a way to say that it must be the wrong action to perform in that situation. An agent may be blameless for doing something that she should not do, but she can be blameworthy only for doing something that she should not do, that is, something morally undesirable. Thus, it seems that, in the example, squeezing the trigger could be wrong regardless of whether Jones performs it on her own, or controlled by Dark Demon. Dark Demon did not intervene to change Jones' intention, hence squeezing the trigger is not only wrong, but it is also blameworthy. Jones deserves blame for performing it, while in a scenario where Dark Demon makes Jones perform the action,

119
that may not imply 'can'.

See Sinnott-Armstrong (1984) and Streumer (2003) about senses of 'ought'

blameworthiness may be appropriately attributed to Dark Demon. It might still seem that it is unfair to blame someone for an act when they could not have done otherwise. What I am suggesting here is not that it is never unfair to blame in such cases. I am suggesting that when it *is* unfair, this is not because the action is not wrong (if we see wrongness as moral undesirability).¹²⁰

In this section I have taken into account a Frankfurt-type case discussed by Haji in which the agent believes that what she is doing is wrong. In addition, the agent cannot act otherwise. Haji claims that the agent is blameworthy even though the action is not wrong, because it is morally wrong for an agent to perform an action only if it is within their power not to perform it. I have suggested that the action in the Frankfurt-type case *can* be seen as wrong (in the sense that it is morally undesirable) and that Haji has then failed to show that blameworthiness does not require wrongness.

1.2

In this section I will consider two opposite cases examined by Haji to suggest that there is no affiliation not only between blameworthiness and wrongness, but also between praiseworthiness and the rightness of the action. In the first case an agent performs a wrong action believing that what she is doing is right. In the second case the agent performs a right action believing that what she is doing is wrong. I will criticise Haji's assessment concerning the agent's blameworthiness and praiseworthiness in these cases. I will conclude the section by presenting Haji's

¹²⁰ Some explanation of 'unfairness' may be called-for. I am not going to provide one here.

positive proposal for a necessary epistemic condition of moral responsibility. I will then contrast it with my own proposal.

Haji discusses a case he names *Howser's Hell*.¹²¹

(...) doctor Howser, attempting to do everything she can to cure her patient, follows the course of treatment which all evidence then available to her indicates will result in a cure: she injects the patient with medicine A. Through no fault of any party, though, the diagnosis is incorrect. Administering A is the wrong treatment and it proves lethal. Had the good doctor only known better, she would have administered medicine B which would have effectively healed the patient.¹²²

Haji notes that this is a case in which the doctor does something wrong - *i.e.* killing the patient. However, Haji claims that the doctor is not blameworthy for giving A. If anything, she might be praiseworthy for doing so. And this suggests that there is no affiliation between blameworthiness and wrongness, and between praiseworthiness and rightness.

I only partially agree with Haji's assessment of the case. I agree that the doctor is blameless for injecting the patient with medicine A and for killing the patient. The reason is that, although those actions are wrong, the doctor acted according to her better judgement that it was required to inject the patient with medicine A. This means that injecting the patient with medicine A and killing the patient are actions that do not possess the feature required to make the doctor blameworthy for performing them, that is its being performed against the agent's

121

This type of case can be found in Feldman (1986).

122

Ibid., p. 528.

better judgement. However, I disagree with Haji's assessment when he claims that the doctor may be rather praiseworthy for her action. There is no reason to believe that the doctor deserves praise for injecting a medicine that has killed a patient. After all, injecting the patient with medicine A and killing the patient are wrong actions, and that is exactly why we would not take the doctor to be praiseworthy for performing them.

The example seems to suggest a consequentialist framework, that whether an action is right or wrong depends on its consequences. Now suppose that consequentialism is true. If an action brings about morally worse consequences than those which the alternative option would bring about, it will be inappropriate to praise the agent for performing the action that brings about worse consequences.¹²³ The reason is that we do not want the action with worse consequences to take place, that is, to be performed, by anyone, in that situation. In *Howser's Hell* there are two alternative options which are mentioned: giving medicine A, which is wrong, and giving medicine B which is right. And it is inappropriate to praise the doctor for performing the wrong action, since it seems inappropriate to praise someone for performing an action that should *not* be performed.¹²⁴ Still, we may feel that the doctor deserves some praise, and indeed I think she does, but *not* for administering the medicine that killed the patient. She rather deserves our praise for *the fact that* she acted according to her better judgement that the right thing to do was injecting A.

123 An exception to this is the following: if I perform a required action instead of a supererogatory action, I might still be praiseworthy for it even if it is not the *best* action I could perform.

124 I am inclined to think that distinguishing between 'objective' and 'subjective' wrongness would unnecessarily complicate my account. I will instead assume that we cannot praise an agent for an action that is objectively wrong even if it is subjectively right. If some praise is deserved in such a case, it is not for the same action. I will say more on this when I discuss Jackson's case in the next paragraph.

As in the case of acting against one's better judgment, also acting according to one's better judgement of what is the right thing to do is a general action. And it is also a praiseworthy-making feature of particular actions that fall under that general action. However, these particular actions are correctly assessed as praiseworthy only if they are also right (or supererogatory) in a non-perspectival sense. In other words, the agent will be praiseworthy for performing a particular action only if that action is right (or supererogatory). This implies that there is then an affiliation between praiseworthiness and rightness of the action, contrary to what Haji claims.

Another case that might suggest that an agent may be praiseworthy for doing something wrong is one presented by Frank Jackson.¹²⁵ In Jackson's case a doctor knows that medicine A will cure the patient partially and that one of B and C will cure the patient completely while the other will be lethal. On the one hand, the right thing to do seem to give the medicine that cures completely, which means that the act of giving A is morally wrong. On the other hand, it seems that the doctor should play safe and give A, to avoid the risk of giving the lethal medicine to the patient. Thus, as we assess the case, we could find ourselves in a situation where we feel that the doctor deserves praise for doing something morally wrong (giving A).

First of all, we should note that the doctor cannot decide to give the medicine that cures completely *the same way* she can decide to give the medicine that cures partially, because she does not know which medicine is the complete cure. Thus, the doctor seems to have three options: she can either (1) give the medicine that cures partially *while knowing* that it is the partial cure; (2) give the medicine that cures completely without knowing that it is the complete cure; or (3) give the lethal

medicine without knowing that it is the lethal medicine. However, from the agent's perspective, these three options boil down to a choice between giving a medicine knowing that it will cure partially and doing something risky, that is giving a medicine having 50% chance that it is the complete cure and 50% chance that it is lethal.

Although it might seem that the agent in this case could deserve praise for doing something wrong (giving the medicine that cures only partially), if we look carefully at the case we realise that we need to make further assumptions to decide whether the doctor deserves praise for giving A. First, we need to establish which option between giving A and performing the risky action is in fact right. This will depend on what moral theory is the true one. Second, we need to consider whether the doctor believes that giving A is right. If giving A is in fact the right thing to do and the doctor believes that giving A is the right thing to do, she may deserve praise for giving A. But if going for the risky alternative is in fact what she should do, and the doctor believes she should instead give A, she will not deserve praise for giving A, because that is in fact the wrong choice to make. Therefore, although Jackson's case may appear to suggest that an agent can be praiseworthy for performing a wrong action, this is not the case since in order to establish whether the an agent is praiseworthy in such a case, we still need to make assumptions about the moral status of the actions among which the agent must choose.

Haji then presents a case that is, in some respect, opposite to *Howser's Hell*, naming it *Deadly's Defeat*:

(...) doing the best she can to murder her patient, doctor Deadly does what credible evidence to which she has access indicates will kill the patient – she injects the patient with medicine C. But once again, this time happily for the patient, there has been an innocent error in diagnosis. Contrary to what Deadly believes, her patient is suffering from a disease which can be cured only by taking C. Giving C results in the lucky patient's full recovery.¹²⁶

Haji claims that, although it is required for Deadly to give C, she is blameworthy for it. I disagree with Haji's assessment of the case. I believe that Deadly is blameless for injecting the patient with medicine C exactly because that is what she was required to do. Injecting the patient with medicine C can be seen as the morally desirable action to perform in those circumstances, and since it is an action that we want to be performed, it would seem senseless to blame the agent *for performing it*. After all, that is exactly what we should hope the agent would do in all situations: perform the required action. Still, we may feel that Deadly deserves some blame. However, we must be accurate in identifying the proper object of blame.

Deadly might deserve blame if she acted against her better judgement. In this respect, unfortunately, the case is under-described. Suppose that Deadly acts believing that injecting her patient with C – and thus killing the patient – is wrong. It seems to me that in this case Deadly is indeed blameworthy. But she deserves blame only for *the fact that* she acted against her better judgement that injecting her patient with C is wrong, not for injecting her patient with C.

Now suppose that she believes that injecting her patient with C – and thus killing the patient – is a morally permissible action. If that is the case she is neither

praiseworthy nor blameworthy for *the fact that* she acted according to her better judgement that injecting her patient with C is permissible. The reason for this is that there is nothing in her better judgment that indicates that she should not perform that action, and giving C is, to her, an action with no obligation attached to it.

Finally suppose that Deadly believes that injecting her patient with C – and thus killing the patient – is required. In this case I am inclined to think that she is praiseworthy for *the fact that* she acted according to her better judgement that killing her patient by injecting C is required. After all, as agents, our knowledge is always limited and even when we do not know what we should do, all we can do is to try to do the right thing. And it seems to me that trying to do the right thing by itself deserves some praise. Still, Deadly is not praiseworthy for injecting the patient with C and for curing the patient, because the reason why she believes it is required to inject the patient with C is not that it will cure the patient. The reason is instead that injecting C will kill the patient. This missing cognitive connection undermines praiseworthiness for having performed the required action.

I then suggest, contrary to Haji's claim, that moral blameworthiness and praiseworthiness are not only tied to belief in what is wrong and required, but also to what is in fact wrong and required in a non-perspectival sense. It seems inappropriate to praise for a wrong action – since we do not want it to happen in the world – as well as to blame for a required action, since it is an action that we think should be performed.

In conclusion, Haji's proposal for a necessary epistemic condition of moral responsibility is the following:

An agent S is morally blameworthy [praiseworthy] for performing action A only if S has the belief that it is wrong [obligatory or right] for her to do A and this belief plays an appropriate role in S's A-ing.¹²⁷

As has hopefully emerged from my analysis, I would reformulate this by adding another condition, which Haji would not accept, about the moral status of the action. My alternative proposal is then that an agent S is blameworthy [praiseworthy] for performing action A only if: (1) S has the belief that it is wrong [right or supererogatory] for her to do A and this belief plays an appropriate role in S's A-ing;¹²⁸ and (2) A is wrong [right or supererogatory] in a non-perspectival sense.¹²⁹

In this section I have taken into account two opposite cases presented by Haji. In the first case, the agent performs a wrong action believing that what she is doing is the right thing to do. Haji claims that the agent is not blameworthy for her action. She might even be praiseworthy. I have criticised Haji's assessment of the case by pointing out that Haji is misplacing praiseworthiness. The agent cannot be praiseworthy for performing a wrong action. She can only be praiseworthy for *the fact that* she acted according to her better judgement. In the second case, the agent seems to perform a right action believing that what she is doing is wrong (although I have pointed out that the case is under-described). Haji claims that the agent is blameworthy for her action. I have criticised Haji's assessment of the case and argued that Haji is misplacing blameworthiness and praiseworthiness. The agent cannot be blameworthy for a right action. She may then be blameworthy only for *the fact*

127

Ibid., p. 529.

128

role in S's A-ing.

I am not discussing here in what sense S's belief must play an appropriate

129

I have changed Haji's 'obligatory or right' to 'right or supererogatory'. However, I do not see this change as reflecting any significant difference between our views.

that she acted against her better judgement that what she was doing was wrong. My response to Haji suggests that there is indeed affiliation between blameworthiness and wrongness, and between praiseworthiness and the rightness of the action. I have concluded the section by contrasting my own proposal for a necessary epistemic condition of moral responsibility to Haji's. My suggestion is that the action must not only be believed to be wrong/right (or supererogatory) by the agent, it must also be in fact wrong/right (or supererogatory) in a non-perspectival sense.

2.0

In this section I will consider three jointly inconsistent propositions on the relation between blameworthiness, wrongness and the agent's mental states. I will then discuss Andrew C. Khoury's rejection of one of those claims in order to resolve the inconsistency, namely the one that states that wrongness is required for blameworthiness. I will then consider a case presented by Khoury and consisting of two scenarios. In one scenario the agent intends to perform a wrong action and is successful in carrying it out. In the second scenario the same agent intends to perform the same wrong action but is unlucky and thus unsuccessful in carrying it out. I will criticise Khoury's assessment of the case in terms of the agent's blameworthiness. I will also reject Khoury's claim that if we accept the proposition that wrongness is required for blameworthiness we must give up the distinction between act-evaluations and agent-evaluations.

In his paper "Blameworthiness and Wrongness", Andrew C. Khoury

considers the following three propositions:

- (1) An agent is blameworthy for some action only if the action is wrong.
- (2) Factors external to the mental states of the agent¹³⁰ at the time of action can affect whether the action is wrong.
- (3) Factors external to the mental states of the agent at the time of action cannot affect whether the agent is blameworthy for the action.¹³¹

Khoury observes that these propositions, if taken together, are inconsistent. He thinks that to resolve this inconsistency the best solution is to give up (1). I disagree with Khoury and want to suggest that (1) and (2) are the true claims, given that Khoury's defence of (3) is not successful. In order to defend (3), Khoury considers two scenarios:

In the first scenario, an assassin carefully aims his gun at his target and pulls the trigger. The victim is killed. In the second scenario, an assassin carefully aims his gun and pulls the trigger, but a bird flies into the path of the bullet before it reaches the intended victim. (...) the two scenarios are indistinguishable from the point of view of each assassin up until the point at which the bullet has left the barrel.¹³²

Khoury claims that the blameworthiness of the two assassins is the same because the mental states of the two assassins are the same at the time they pull the

130 A factor external to the mental states of an agent is something that changes without there being a change in the mental states of the agent.

131 Khoury (2011), p. 135.

132 *Ibid.*, p. 140.

trigger.

However, it seems to me that the focus of Khoury's assessment is ambiguous. There is a sense in which the blameworthiness of the two assassins is obviously not the same. In the first scenario, the assassin can be blamed not only for pulling the trigger but also for killing the victim. In the second scenario, the assassin can be blamed only for pulling the trigger for the bullet does not reach the victim. Hence, in a way, the presence of the bird changes *what* the assassin is blameworthy for. Now if Khoury's aim is to defend (3), then what he must be concerned with is only *whether* the assassin is blameworthy for pulling the trigger. If this is the case, then it is true that the blameworthiness of the two assassins may be the same. However, I disagree with Khoury that the assassins' blameworthiness is the same *only* because of their mental states at the time they pull the trigger. In order to be blameworthy, both assassins must believe that pulling the trigger as they aim at the victim is wrong. For instance, they may believe that pulling the trigger while aiming at the victim is a morally risky action. They may believe that pulling the trigger might kill the victim and that that is wrong. However, if the action were in fact not risky, and not wrong to perform, we would not blame the assassin for pulling the trigger while aiming at the victim. For example, certainly we would not blame the assassin if he aimed at the ground, or for aiming at a shooting target during his training. Therefore, in order for the assassins to be blameworthy for pulling the trigger while aiming at the victim, they must believe that this action is wrong *and* the action must be wrong in a non-perspectival sense (*e.g.* because it is risky as it may cause the death of the victim). It seems then that there are

factors external to the agent's mental states at the time of the action – namely the wrongness of the action – that will affect blameworthiness.

Khoury then moves to his defence of claim (2), referring again to the example of the assassin:

Let us imagine that the target in both cases was a Mother Teresa type figure who helps people in need and does a great deal of good for the world. In the [scenario] (...) where the assassination is successful, the world suffers a great loss and utility is not maximized. For this reason, the action would be wrong on utilitarian grounds. But in the second case, there was no assassination and so the action cannot be wrong for the same reasons. Furthermore, let us suppose that the bird that flew into the path of the bullet was actually carrying a new and extremely contagious and dangerous strain of bird flu. Had the bird avoided the bullet, it would have infected other birds which in turn would have infected humans and caused a global pandemic. (...) we might judge the action to be right on utilitarian grounds [since] (...) utility was maximized by shooting the bird.¹³³

I agree with Khoury that this gives some reason to think that (2) is true. Khoury later claims that rejecting (2) or/and (3) forces us to give up the distinction between agent-evaluations and act-evaluations. I disagree with this claim. Khoury speaks of agent-evaluations as those evaluations concerned with internal features of the agent. He instead refers to act-evaluations as those evaluations concerned with external factors. I disagree with Khoury's claim because it seems to me that by rejecting (3) and endorsing (1) we would not give up this distinction. We attribute blameworthiness *to an agent for an action*. If you endorse my view that an agent is blameworthy for an action only if she acts believing that what she is doing is wrong and if the action is wrong in a non-perspectival sense, then the wrongness of the action

is exactly that external factor that affects the agent's blameworthiness. We can both evaluate an agent's internal feature without blaming the agent for having that feature (*e.g.* if she is not aware of it or/and she has no control over it) and evaluate an act without blaming the agent for performing it (because an act can be wrong and blameless). Hence, we can allow the distinction between agent-evaluation and act-evaluation without rejecting the claim that blameworthiness requires wrongness.

In this section I have taken into account three jointly inconsistent propositions on the relation between blameworthiness, wrongness and the agent's mental states. I have criticised Khoury's assessment of a case utilised to defend claim (3) and reject (1). The case includes two scenarios: in the first scenario the agent is successful in carrying out a wrong action he intended to perform. In the second scenario the same agent is unlucky and unsuccessful in carrying out the wrong action he intended to perform. Khoury claims that in both scenarios blameworthiness is the same. I have undermined this claim and suggested that the wrongness of the action is a factor, external to the agent's mental states at the time of the action, that affects what the agent is blameworthy for. I have therefore suggested that we should maintain (1), and rather reject (3). Finally since my proposal allows us to consider both an action's moral status and the agent's blameworthiness for that action, I have rejected Khoury's claim that if we accept (1) we must give up the distinction between act-evaluations and agent-evaluations.

2.1

In this final section I will take into consideration Khoury's epistemic condition of blameworthiness. He claims that an agent is blameworthy for an action only if the agent

ought to believe that that action is wrong. Khoury wants to show that this claim is correct and consistent with claims (2) and (3). According to Khoury, an agent ought to believe that her action is wrong either in virtue of her actual beliefs about her action or in virtue of the beliefs about her action which she should have. I will consider some examples presented by Khoury to support his claim and show that in these cases – which Khoury assesses as blameworthy – there is at least one sense in which it is not the case that the agent ought to believe that what she is doing is wrong. I will also discuss Khoury's interpretation of “ought to believe” in terms of reasonable moral expectations on the agent.

Khoury points out that in order to understand the relation between blameworthiness and wrongness we need to take into account both epistemic and normative considerations. Thus, Khoury considers a claim consistent both with (2) and (3): “an agent is blameworthy for some action only if she ought to believe the action to be wrong”.¹³⁴ Khoury explains that what the agent ought to believe – in this sense – depends on the agent's mental states, on how the agent views the world. Khoury argues that what an agent ought to believe regarding the normative status of her action can be determined by her actual beliefs about what she is doing and by the beliefs about what she is doing that she ought to have, *considering her mental states*. Consequently, according to Khoury “an agent is blameworthy only if she ought to believe that her action is wrong in virtue of either her actual beliefs about her action or the beliefs about the action that she ought to have”.¹³⁵

Let us take the agent's actual beliefs first. Khoury's idea is that if I have an actual belief, for instance, that I am stepping on a baby, I ought to think I am doing wrong. If Hitler had the

¹³⁴ Khoury (2011), p. 142.

¹³⁵ *Ibid.*, p. 144.

actual belief that he was ordering the mass extermination of Jews, he ought to have believed that mass extermination is wrong. These conditionals are true in virtue of the content of the agents' actual beliefs. And, according to Khoury, this is, in part, why I am blameworthy for stepping on a baby even if I do not believe I am doing wrong, and why Hitler is blameworthy for ordering the mass extermination of Jews even if he did not believe he was doing wrong.

I find Khoury's proposal puzzling. It seems at least plausible that the actual beliefs of an agent about what she is doing may not entail that she ought to think, *from her point of view*, that the action is wrong. There is nothing in the content of the belief that I am stepping on a baby, by itself, that suggests that I should believe, from my point of view, that it is wrong to do it. The reason is that it might be the case that I have most (internal) reason to think that my action is required if we imagine, for instance, a situation where I believe that I should step on a baby in order to save the life of two other babies. In this case it is not true that, from my point of view, I should think that stepping on that baby is wrong. Therefore, the reason for being blameworthy for stepping on a baby is not that I should believe it is wrong. Hitler was most likely aware that he was ordering the mass extermination of Jews but, at the same time, he might also have thought that that was something morally required of him to do. And he might have thought so for certain other misguided beliefs and mental states he had. Thus, from his mental states it does not follow that he ought to think that ordering the mass extermination of Jews is wrong.

Now let us consider the beliefs an agent ought to have about what she is doing, as opposed to the agent's actual beliefs. Khoury examines the following case:

Let us suppose that Jim is babysitting Alexa, a five year old whose mother has told Jim that Alexa

is deathly allergic to peanuts and has given Jim a note explaining this along with emergency contacts. However, Jim's attention is elsewhere. He is thinking about the football game that is about to begin. Alexa's mother leaves and Jim turns on the game. A couple of hours later Alexa wants lunch. Jim, preoccupied with the game, makes a peanut butter and jelly sandwich and gives it to Alexa, who dies of an allergic reaction.¹³⁶

Jim did not think that what he was doing was wrong. Still, according to Khoury, Jim ought to have thought so. The fact that he did not think that what he was doing was wrong is a fact about the quality of his will – Khoury states. And it is a fact about his mental states. A fact that makes, at least in part, Jim blameworthy.

I disagree with Khoury's assessment of the case. Generally I think that, as the example is presented, it is not clear what type of case we are dealing with. However, I will assume that the most plausible reconstruction of the example is the following: Alexa's mother told Jim about the child's allergy, but Jim was thinking about the game that was about to start, so he did not hear what Alexa's mother was telling him. Or, to be more accurate, he heard Alexa's mother's words, but he failed to store that information so that it could play a role in his future action. There is indeed a sense in which he should have thought that giving peanut butter to Alexa was wrong. This is quite straightforward: he should have thought so because he was told that Alexa was allergic to peanut butter. However, let us remind ourselves that what Khoury is interested in is what the agent should believe *considering his mental states*. Now it seems to me that the mental contents Jim was aware of at the time he gave peanut butter to Alexa did not entail that he should believe that giving Alexa peanut butter

was wrong. At the time he was told that Alexa was allergic to peanut butter he was thinking about the game. Hence, from Jim's point of view, it is not the case that he should believe that giving peanut butter to Alexa is wrong, since he was not aware of the fact that Alexa was allergic to peanut butter. Jim seems then blameless for not thinking that his action is wrong, and consequently also blameless for giving Alexa a peanut butter and jelly sandwich.

Jim could deserve some blame but we must be careful in identifying *what* he would be blameworthy for. For example, Jim may realise that he is not paying attention when Alexa's mother is giving him instructions. He may think that he should ask her to repeat or call her back if he realises it after she has left. If Jim decides that he cannot be bothered to ask Alexa's mother to repeat the instructions because he just wants her to leave so that he can watch the game, then it seems like Jim will be blameworthy for not holding certain beliefs that he would have held if he had asked Alexa's mother to repeat the instructions for him. But this does not entail that he will be blameworthy for not holding the particular belief that giving Alexa peanut butter is wrong, because he did not foresee that he could end up not holding such a belief, and consequently performing such a fatal action.¹³⁷

Khoury thinks that the idea that an agent 'ought to believe' could be understood as 'it is reasonable to expect the agent to believe'.¹³⁸ But since Khoury thinks that these expectations are still dependent on the agent's mental states at the time of the action, it seems that there is a sense (relevant to responsibility) in which it might be unreasonable to expect Jim to think that giving Alexa peanut butter is

¹³⁷
¹³⁸ responsibility.

See chapter 5 for more discussion on this foreseeability condition of
Khoury (2011), p. 144.

wrong, since he does not hold the belief that Alexa is allergic to peanut butter and so he cannot infer that it is wrong to give her peanut butter. It is unreasonable to expect Jim to believe that giving Alexa peanut butter is wrong when he does not have any internal reason to hold that belief.

Khoury claims that if Alexa's mother had not mentioned to Jim that Alexa was allergic to peanut butter Jim's will would not have displayed "negligent lack of concern".¹³⁹ And in this case, according to Khoury, it would be unreasonable to expect him to believe that giving Alexa's peanut butter is wrong. I agree with Khoury that there is a sense in which if Alexa's mother did not mention Alexa's allergy to Jim, from Jim's point of view (*i.e.* his mental states) it is unreasonable to expect him to believe that giving Alexa peanut butter is wrong. But there is also a sense (which I take to be relevant to responsibility) in which it is unreasonable to expect Jim to think that it is wrong to give Alexa peanut butter if Jim does not hold the belief that Alexa is allergic to peanut butter. This is because lack of concern on its own is not a decisive factor in appropriately attributing blameworthiness. Lack of concern can be certainly identified as a feature of a bad moral agent, but not necessarily of a blameworthy agent. The reason for this, in my view, is that what makes an agent blameworthy is, in the end, the decision made against their better judgement, regardless of their concern. You may lack concern for strangers' health but you could still decide to aid a stranger if you come across someone in need. You might do it because you think it is the right thing to do, without this meaning that you are concerned with that person's health. And it seems that you would still deserve some praise for doing the right thing. On the other hand, in a situation where you are not

concerned with a stranger's health and you also think that it is *not* required of you to aid that person, what is most relevant to responsibility is the fact that you believe that it is not required of you to aid that person and so, in not helping her, you are acting according to your better judgement. This makes you blameless for not helping that person.

Our better judgement is all we can base our decisions upon in trying to do the morally right thing. In Jim's case, regardless of his lack of concern, all he can do is to decide based on what he knows. And he has no internal reason to think that giving Alexa some peanut butter is wrong, because he does not hold the belief that she is allergic to peanut butter. So it is unreasonable to expect him to believe that it is wrong to give Alexa peanut butter. Consequently, he seems blameless for his decision to give Alexa a peanut butter and jelly sandwich (and for the resulting incident).

Khoury claims that "when an agent fails to meet a reasonable moral expectation, this may warrant some reaction such as indignation or resentment, or it may simply impair her relationship with other people".¹⁴⁰ I agree with Khoury. Actually, in Khoury's claim we can see the appropriate distinction between blameworthy and blameless agents. When an agent deserves resentment for failing to meet a reasonable moral expectation, it is because she is blameworthy for that failure. And whether the agent deserves resentment depends on whether she believed that what she was doing was wrong when she failed to meet the moral expectation. On the other hand, when the agent's failure to meet a reasonable moral expectation simply impairs her relationship with other people, it

is because the agent is blameless for that failure. And she may be blameless because she did not believe that what she was doing was wrong when she failed to meet that moral expectation.

In conclusion, Khoury's proposal is that an agent is blameworthy for some action only if it is reasonable to expect the agent to think her action is wrong. And this is a claim consistent with (2) and (3).¹⁴¹ Therefore, Khoury thinks that claim (1) should be rejected to resolve the inconsistency between (1), (2) and (3). My view is instead that an agent is blameworthy for some action only if she believed it was wrong and the action was wrong in a non-perspectival sense. This claim is consistent both with (1) and (2). I therefore think that (3) should be rejected as false, since the moral status of an action, which is something external to the agent's mental states, *can* affect the agent's blameworthiness.

In this final section I have considered Khoury's epistemic condition of blameworthiness. Khoury's claim is that an agent is blameworthy for an action only if the agent *ought to believe* that the action is wrong – a claim which he sees as consistent with (2) and (3). According to Khoury, an agent ought to believe that her action is wrong either in virtue of her actual beliefs about her action or in virtue of the beliefs about her action which she should have. First I have shown that the examples presented by Khoury to support his view can be seen as cases of blameless conduct. Moreover, I have suggested that in those cases there is at least one sense in which it is not the case that the agent ought to believe that what she is doing is wrong, from the agent's point of view. I have also discussed Khoury's interpretation of “ought to believe” in terms of reasonable moral expectations on the agent. Again,

I have suggested that there is at least one sense in which, in the cases presented by Khoury, it might be unreasonable to expect the agent to believe that what she is doing is wrong. I have then concluded that Khoury's argument to support his epistemic condition of blameworthiness (which is compatible with (2) and (3)) is not convincing. I think (3) should be rejected instead. My proposal, which is compatible with (1) and (2), still allows us to resolve the inconsistency between the three claims. This means that we should accept (1) as true, that is the claim that an agent is blameworthy for some action only if the action is wrong (*i.e.* the *objective view*).

3.0

In conclusion, in this chapter I have defended the *objective view* from some objections by Haji and Khoury. I have also suggested a view about the relation between blameworthiness and wrongness. My view combines two necessary conditions of blameworthiness as follows: an agent is blameworthy for performing an action only if (a) she believes the action is wrong *and* (b) the action is wrong in a non-perspectival sense.

In 1.0 I have considered two versions of a case presented by Haji. In the first version, the agent performs a wrong action and believes that what he is doing is morally wrong. In the second case the agent performs a permissible action and believes that what he is doing is wrong. I have criticised Haji's assessment of the case by pointing out that Haji misplaces the agent's blameworthiness. In the version of the case where the agent performs a wrong action, he is blameworthy for it. In the version of the case where the agent performs a permissible action, he is blameless for performing the action and blameworthy only for *the fact that* he

acted against his better judgement that he should not perform that action. This indicates that blameworthiness is not only associated with the agent's belief that she is performing a wrong action – as Haji claims – but also with the wrongness of the action.

In 1.1 I have taken into account a Frankfurt-type case, discussed by Haji, in which the agent believes that what she is doing is wrong but she cannot act otherwise. Haji claims that the agent is blameworthy but the action is not wrong, because it is morally wrong for an agent to perform an action only if it is within their power not to perform it. I have rejected this claim and suggested that the action in the Frankfurt-type case *can* be seen as wrong and that Haji has failed to show that blameworthiness does not require wrongness.

In 1.2 I have taken into account two opposite cases presented by Haji. I have criticised Haji's assessment of the first case by pointing out that Haji is misplacing praiseworthiness. The agent cannot be praiseworthy for performing a wrong action. She can only be praiseworthy for *the fact that* she acted according to her better judgement. I have criticised Haji's assessment of the second case and pointed out that here Haji is misplacing blameworthiness. The agent cannot be blameworthy for a right action. She can be blameworthy only for *the fact that* she acted against her better judgement. My response to Haji suggests that there is indeed affiliation between blameworthiness and wrongness, and between praiseworthiness and rightness of the action.

In 2.0 I have taken into consideration three jointly inconsistent propositions on the relation between blameworthiness, wrongness and the agent's mental states. I have criticised Khoury's assessment of a case utilised to defend claim (3) and reject (1). I have shown that the case does not suggest that factors external to the agent's mental states at the time of the action do not affect blameworthiness. I have therefore suggested that we should

rather maintain (1) and reject (3). I have also refuted Khoury's claim that if we accept (1) we must give up the distinction between act-evaluations and agent-evaluations.

Finally, in 2.1, I have considered Khoury's claim that an agent is blameworthy for an action only if the agent *ought to believe* that the action is wrong – an epistemic condition of blameworthiness consistent with (2) and (3). According to Khoury, an agent ought to believe that her action is wrong either in virtue of her actual beliefs about her action or in virtue of the beliefs about her action which she should have. First I have shown that the examples presented by Khoury to support his view can be seen as cases where there is at least one sense in which it is not the case that the agent ought to believe that what she is doing is wrong, from the agent's point of view. I have also discussed Khoury's interpretation of “ought to believe” in terms of reasonable moral expectations and suggested that there is, again, at least one sense in which, in the cases presented by Khoury, it might be unreasonable to expect the agent to believe that what she is doing is wrong. I have then concluded that Khoury's argument for his epistemic condition of blameworthiness – which is compatible with (2) and (3) – is not convincing. We should rather reject (3) and accept my proposal, which is compatible with (1) and (2) and still allows us to resolve the inconsistency between the three claims. We should then accept the claim that an agent is blameworthy for some action only if the action is wrong, which means we should accept the *objective view*.

Chapter 4 – Transferring Blame: Factual and Moral Ignorance

In this chapter I will undermine some objections to the thesis that whenever an agent acts from ignorance, she is culpable for the act only if she is culpable for the ignorance from which she acts (Ignorance Thesis). Volitionists support this thesis and claim that ignorance is culpable when it is the result of an earlier culpable action, while objections mainly come from attributionists who hold that in order to explain responsibility for ignorant behaviour we do not need to trace back to culpable actions that caused that ignorance. Here I will focus not only on factual ignorance, but also on moral ignorance.

In the first three sections I will consider factual ignorance. I will look at cases which Matt King takes to be counterexamples to the thesis that in order to explain responsibility for ignorant behaviour it is necessary to trace it back to responsibility for that ignorance. In 1.0 I will discuss a case in which the agent enacts a plan that includes an action of which the agent will be ignorant when he performs it. In 1.1 I will examine some actions particularly extended in time which include some shorter actions of which the agent is unaware when he performs them. In 1.2 I will take into account a case of negligence where an agent brings about harm out of supposed culpable ignorance. I will disagree with King's evaluation of the cases in terms of the agent's blameworthiness, and I will dispute that such cases support the claim that in order to explain responsibility for ignorant behaviour we do not need to consider whether the agent's ignorance was itself culpable.

In the last three sections I will focus on moral ignorance and I will defend the

Moral Ignorance Thesis from some objections raised by Alexander Guerrero. In 2.0 I will illustrate the Moral Ignorance Thesis. This is the thesis that whenever an agent acts from moral ignorance, she is culpable for the act only if she is culpable for the ignorance from which she acts. I will first present Guerrero's definition of something done from ignorance, then his identification of different ways in which an agent can be ignorant of a fact, and finally his view of culpable ignorance. I will claim that Guerrero's conditions of culpable ignorance are necessary but not sufficient, as he seems to claim, and I will suggest an additional condition. In 2.1 I will take into consideration the widely debated case of the Ancient Slaveholder presented by Gideon Rosen. I will undermine Guerrero's attempt to show that cases such as the Ancient Slaveholder do not support the Moral Ignorance Thesis. Finally, in 2.2 I will discuss a case which Guerrero takes to be a counterexample to the Moral Ignorance Thesis, since it is meant to be a case in which the agent acts *from* moral ignorance, he is *not* culpable for that ignorance but he *is* culpable for his action. I will show that the case presented by Guerrero is not a case of something done from ignorance. I will then suggest that Guerrero's definition of something done from ignorance does not allow us to assess the case correctly and I will propose a way to amend the definition by adding a further condition.

1.0

In "Epistemic Tracing" Matt King argues for the claim that tracing is not necessary to explain responsibility for ignorant behaviour.¹⁴² In other words, according to King,

142

See also Vargas (2005) for a similar view.

it is *not* the case that we are responsible for an action performed from ignorance only if we are responsible for an earlier action that caused that ignorance.¹⁴³ The aim of this section is to undermine King's argument for this claim. King considers the following case of factual ignorance:

(...) Benson doesn't know that the button he pushes to start his coffee maker is hooked up to a bomb (...), he set it up last night so that it would detonate the bomb when he made his morning coffee. (...) [But] he also took a pill shortly thereafter, which temporarily rendered his memories regarding the button inaccessible. He subsequently forgot about his plan to blow up the bakery, about planting the explosive, wiring the coffee maker, etc. As a result he intended to erase the relevant epistemic states, such that when he presses the button in the morning, he no longer knows that he is not just making coffee, but setting off a bomb.¹⁴⁴

King takes Benson to be blameworthy for the bomb's destruction. King thinks that this is a counterexample to the claim that tracing is necessary to explain responsibility for ignorant behaviour. King claims that the reason why Benson is blameworthy for the bomb's destruction is not that he is responsible for taking the pill. What makes Benson blameworthy for the explosion is that when he switches on the coffee maker, although he is not aware that he is causing an explosion, he detonates the bomb *knowingly*. It is done knowingly because Benson enacts a plan and performs a number of actions to realise his plan (*e.g.* wiring elements together, packaging the explosive, etc...). And after all his plan is to blow up the bakery *in*

143

The opposite view is endorsed by Fischer & Tognazzini (2009).

144

Matt King (manuscript), p. 2. King stipulates that Benson is still pleased with the explosion of the bakery although he doesn't know his role in it.

such a way that he will not be aware when he causes the explosion. King's claim is then that what explains that Benson is responsible for the explosion is not the fact that he is responsible for taking the pill, and this is a counterexample to the claim that in order to explain responsibility for ignorant behaviour we need to trace back to an earlier responsible non-ignorant action that caused that ignorance.

It seems to me that if we consider Benson's plan as one single action, then it is true that the action has been performed knowingly. However, Benson's particular action of starting the coffee maker is still done in ignorance of the action's implications. When Benson starts the coffee maker, he does not know about his plan *then*. There is nothing that suggests to him that if he starts the coffee maker he will blow up the bakery. And if he had not taken the pill, he might have refrained from starting the coffee maker. He might have changed his mind about his evil plan (maybe this is exactly the reason why he took the pill). Benson knows nothing about his plan in the morning. He only knows that starting the coffee maker will produce some coffee as an outcome. Therefore, a way to explain his responsibility for blowing up the bakery is indeed to trace it back to taking the pill, knowing that this will erase his memory temporarily and will result in his starting the coffee unaware of what that action will bring about.

In order to explain how Benson's case can be seen as a counterexample to the thesis that tracing is necessary to explain responsibility for ignorant behaviour, King gives two examples of action performed using technology:

I can set up my email to send out a message at a particular time. I haven't emailed anyone

until that time, and it is possible that when I send it I will be unaware that I am sending it. Perhaps I forgot about it altogether. (...) In fact, my tendency to forget about having to send the email is precisely why I scheduled it to happen in the first place.¹⁴⁵

King claims that it would be wrong to say that I emailed under ignorance. However, I think there might be an issue regarding the individuation of the actions in this case. King might be claiming that I perform a single action of emailing that starts when I set up the system and continues until the email is sent. Or he might be claiming that I perform two actions: setting up the email and sending it later. But it seems to me that a less confusing way of individuating the actions in this case would be to say that I perform one action, that is setting up my email, and when the email is sent later this is only a *result* of my earlier action. Although I'm open-minded about how other actions should be described, according to my description of this case I knowingly set up my email and that is what matters for my being responsible for the result (*i.e.* the email being sent), regardless of whether I am aware that the email is being sent when that happens.

The other example King mentions is the following: "I routinely set my dishwasher to start on an hour delay, at least in part because of my tendency to forget to start it otherwise".¹⁴⁶ King asserts that in this case I run the dishwasher intentionally, regardless of my occurrent mental states when its cycle begins. King claims that the same applies to Benson's case: Benson acted knowingly both in taking the pill and in starting the coffee maker because he enacted his plan fully aware of what he was doing. The principle King is suggesting here might be that if a

145

Ibid., p. 3.

146

Ibidem.

plan is fully aware, all the smaller actions of the plan are not in ignorance even if the agent is not aware of the plan at the time of the smaller actions.

However, I think that King is mistaken since the parallel drawn between Benson's example and the technology examples does not support King's case. Actually, there is an important difference between Benson's case and the email or the dishwasher example. Benson performs not one but at least two relevant intentional actions that result in the explosion of the bomb: he sets up the explosive *and* he pushes the button of the coffee maker. Both actions are necessary for the explosion to take place. He performs the former intending to blow up the bakery and the latter intending to make coffee. It seems to me that he is blameworthy for pushing the button of the coffee maker exactly because he is blameworthy for taking the pill in accordance with his plan.

Now suppose that Benson believes he will easily carry out his plan the morning after, so he does not take any pill. Still, his colleague does not trust him so he sneakily drops the pill into Benson's drink. In the morning, Benson will not be blameworthy for blowing up the bakery, since he pushes the button believing that by doing that he will only make coffee. In contrast, in the case of the email, for instance, in order to send out a message I only need to set up my email and I don't (need to) perform any other later action that will affect the result (the message being sent out). Thus, in a way the action of sending out the message is already contained in the action of setting up my email without any other intervention on my part. It is then irrelevant whether I am aware that the message is being sent out at the time that happens.

The same applies to the case of the dishwasher. I set the dishwasher to start in one hour. I believe that the dishwasher will start in one hour. I do not need to do anything else to start the cycle. And it is irrelevant whether I am thinking about the fact that the cycle is beginning when that happens, one hour later. I am responsible both for setting up the dishwasher and for the result – the fact that the cycle will begin after one hour. Still I might *not* be responsible if the cycle does not begin (*e.g.* because the dishwasher is malfunctioning). Therefore, I am inclined to conclude that the parallel drawn between Benson's example and the technology examples does not support King's claim that tracing is not necessary to explain responsibility for ignorant behaviour.

1.1

In order to strengthen his argument for the claim that tracing is unnecessary to explain responsibility for ignorant behaviour, King considers some actions that are particularly extended in time: taking a walk and baking a cake.

We may interrupt such activities with other actions without ceasing the original activity altogether. If one stops to smell the roses, one has not thereby stopped taking a walk. Or one may be in the middle of baking a cake and have to run out to the grocery store for a missing ingredient. It is accurate when seeking assistance in the store to say, "Well, I'm baking a cake and need vanilla".¹⁴⁷

King claims that the organisation of our mental states, that is the persisting intentions and beliefs we have about our actions, are what makes a difference between a case in which I stop my action to do something else and a case in which I don't.

[Suppose that] Sergio (...) consider[s] the possibility that he stops to smell the roses during his walk. He proceeds anyhow. When he stops to smell the roses, he may do so absent-mindedly, thinking instead only of the complicated philosophical argument he's trying to work out in his head.¹⁴⁸

King claims that if the action of smelling the roses were unjustified, it would count as reckless. Hence, this means that Sergio performs it – more or less – knowingly.

However, it seems to me that, given how King's examples are presented, it is not clear that they support his case. Leaving aside the question of whether the action would be correctly defined as reckless, let us imagine that smelling roses during Sergio's walk were something wrong to do. Suppose that Sergio has to meet a friend and plans to take a walk before the appointment. During the walk Sergio considers whether to stop and smell the roses. He does not have enough time, and if he stopped he would arrive late at the appointment. He knows he should not do it, but he prefers to take his time and decides to do it anyway. Sergio then stops and smells the roses *regardless of his better judgement*. This looks like a case in which Sergio is blameworthy for stopping to smell the roses and for arriving late at the appointment.

But this is not a case of ignorant behaviour. During his walk Sergio does not need to continuously think that he should not stop to smell roses in order to be blameworthy for that action. But once he has made his choice, he is responsible for the action that follows, I am inclined to say, at least in normal cases.¹⁴⁹ However, a substantially different case would be one in which Sergio stops to smell the roses *because* he has forgotten about his appointment. This seems to be a case of ignorance in which he is blameless for stopping to smell the roses, unless his forgetting is caused by a prior action of his (as in Benson's case). Thus, I believe that King, in using these examples of actions particularly extended in time, does not offer strong support to his view that tracing is not necessary to explain responsibility. King's idea that I find mistaken is the following: King wants to suggest that for short ignorant actions we would be tempted to say that the responsibility for the action depends on the responsibility for previous actions. But if we look at actions extended in time this is not so, because we have to consider the whole plan, which includes also smaller actions that might be seen as ignorant actions, when they are not. According to King, they are not ignorant actions, but simply actions of which we are not aware when we perform them.

1.2

In order to argue for the claim that tracing is not necessary to explain responsibility for ignorant action, King considers another case in which an agent brings about harm out of supposed culpable ignorance.

¹⁴⁹ See chapter 6 for more discussion on the type of awareness required for moral responsibility.

Suppose a doctor [, Jo,] fails to stay up to date on the latest medical research relevant to her practice. As a result she causes harm to a patient, harm she would have known how not to cause had she properly informed herself. She could have read the relevant recent medical research, but neglected to. (...) suppose that she just doesn't think about the risk of harm at all, and (...) [doesn't] decide to ignore the research. Instead, she (...) simply fail[s] to think of informing herself, though she should.¹⁵⁰

King claims that Jo is negligent and therefore she is responsible for the harm her negligence brings about. She is negligent in causing harm because she should have realised the unjustified risk she was posing to her patient. Staying up to date on the latest medical research is a standing duty of physicians. Jo did not realise that she failed to recognise a risk she ought to recognise. So negligence just traces back to negligence. Therefore, according to King, we don't need to trace back: either that negligence is sufficient to explain responsibility without tracing, or tracing back to a previous instance of negligence will not help.

I disagree with King's evaluation of the doctor's action as far as moral blameworthiness is concerned. Staying up to date is certainly a standing duty of physicians. A doctor who does not stay up to date on the latest medical research is a bad, unreliable doctor, since she doesn't do what her role requires from her. However, from this alone it does not follow that this doctor is also *morally blameworthy* for failing to stay up to date on the latest medical research.¹⁵¹ Surely Jo

150 Matt King (manuscript), pp. 5-6.

151 I realise and accept that an implication of my view is that I can be morally blameless for failing to perform my duty (even when it is not a case in which there is some countervailing duty).

could have read the relevant research. However, she did not do that because she did not realise that there was a risk of harming someone. Now the question is whether it is reasonable to expect Jo to do something to make herself realise that risk. And it seems to me that unless someone reminds her about it or she somehow reminds herself, it is unreasonable to expect her to bring herself to think about something. Expecting this would be like expecting her to have a different mind. We may be disappointed or frustrated that she has not realised about the risk of harm, but it would not be appropriate to resent her for her failure. We may simply conclude that she is blamelessly flawed for failing to stay up to date on the latest medical research, since at no point did she deliberately fail to fulfil this obligation. We may not want to have Jo as our doctor, but it seems inappropriate to blame her for something over which she has no control.

A scenario in which Jo would be blameworthy may look like this: in the past Jo didn't think of reading the latest issue of a particular medical journal. Luckily, that brought no harm. However, she then realised that she could have caused some harm to her patient, so she thought she should set a reminder to check the latest issue of that medical journal as soon as it is out the next time. However, she could not be bothered to set the reminder, and she decided to rely only on her memory (hoping it would not betray her again) although she thought she shouldn't. This is a case in which Jo would be blameworthy for not setting the reminder and, indirectly, for not reading the latest issue of the journal if she forgets to. She would also be blameworthy for harming her patients if the content of the journal suggested that she should operate differently.

King then compares the case of Jo to the case of another doctor who is not

negligent:

Suppose Joaquin had been keeping up with the latest research, but the particular journal with the pertinent information was delayed in the mail (and so a day late). Supposing he treats a patient under identical circumstances as Jo (...).¹⁵²

King takes Joaquin to be blameless for harming the patient. The reason is that Joaquin is not negligent. He takes appropriate precautions against causing harm, but he lacks the opportunity to inform himself. And to explain the difference between Jo and Joaquin we don't need to appeal to tracing – King maintains. The difference is simply that while Jo should have known better, this is not the case for Joaquin. The opportunity that Jo had to inform herself makes her failure to know better negligent, and thus culpable, whereas this is not the case in Joaquin's circumstances. King's claim is then that the question of whether the doctor is responsible for harming the patient is a question of whether the doctor is negligent and not of whether the doctor was culpable for their ignorance.¹⁵³

I agree with King that Jo should have known better. However, it might still be the case that Jo has failed to know better blamelessly. It is true that she did not keep up with the latest medical research although she could. But the relevant question here is why she did not do so. Again, if she did not make any decision that she knew might cause harm to her patient, it is not appropriate to blame her for the harm caused, since, after all, she has no control over what thoughts will pop into her head (*e.g.* thinking of reading the latest issue of a certain medical journal).

Some people may have the intuitions that Jo's case is 'more blameworthy' than

152

Ibid., p. 7.

153

Ibid., p. 8.

Joaquin's case. The reason is that they see Jo's failure as something that is generated within herself, whereas Joaquin's failure is seen as caused by external factors (*i.e.* the delay in the mail). Although I acknowledge this intuition, I think it is misguided since there are several things that are internal to us but over which we do not have control (*e.g.* desires, fears, physical conditions) or of which we are not even aware. The mere fact that they belong to us does not make the attribution of responsibility more appropriate, since it seems to me that we should rather be held responsible for what we do about them once we have become *aware* of them.

King concludes by comparing the case of Jo to the case of a drunk driver. What makes a drunk driver blameworthy – King claims – is not getting drunk, but the fact that he does not take appropriate precautions. It is wrong to drive while drunk not because it is wrong to get drunk, since getting drunk only increases a risk of wrongdoing. According to King, the same applies to the doctors' example. Failing to keep up with medical research increases the chance that the doctor will harm their patients. And it is wrong to neglect to stay up to date on the latest medical research because harming a patient would be wrong, not because not staying up to date on the latest medical research is wrong in itself.¹⁵⁴

It seems to me that this comparison is misleading though. The focus of these cases is the blameworthiness of the agent and not simply the wrongness of the action. I agree that the drunk driver is not doing something wrong merely because he got drunk. What is wrong is the fact that he drives while being drunk, since this is potentially harmful. But in terms of the blameworthiness of the drunk driver, he seems blameworthy only if he knew, when he started drinking, that he would (or might) end up driving his car while drunk, and if he knew that that should not be done because of the risk of harm involved. The driver

seems culpable only if there was a previous episode in which he was clear-headed and he decided to take a risk, knowing that this should not be done. If there is no such episode, the drunk driver may be blameless. If the drunk driver didn't think that he would end up driving drunk, he seems blameless. Or another blameless version of the case is one in which the drunk driver thought he would end up driving, but he didn't think that that would be risky, because, for instance, he sees himself as an excellent driver.

The same applies to the case of the doctor. It is true that what makes 'not staying up to date on the latest medical research' wrong is the fact that it increases the chances of harm to the patients. But since we are talking about blameworthiness, and not merely about the wrongness of the action, we need to consider whether the doctor thought that there was some risk of harm involved in what she was doing. And we also need to consider whether she thought it was a risk that should not be taken. In conclusion, for an agent to be blameworthy it is not sufficient that she fails to fulfil an obligation, since an agent can fail to fulfil an obligation blamelessly or culpably. And the agent is culpable for failing to fulfil an obligation if and only if she thinks that she has that obligation, and she chooses not to fulfil it.

2.0

In his paper "Don't Know, Don't Kill: Moral Ignorance, Culpability and Caution", Alexander A. Guerrero argues against the following thesis:

Moral Ignorance Thesis (MIT) Whenever an agent acts from moral ignorance, he is culpable for the act only if he is culpable for the ignorance from which he acts.¹⁵⁵

The supporters of this view¹⁵⁶ tend to hold the parallel view about factual ignorance. However, the Moral Ignorance Thesis is considered more controversial.¹⁵⁷ The aim of the following sections is to undermine Guerrero's objections to the Moral Ignorance Thesis.

I claim that Guerrero's attack on the Moral Ignorance Thesis depends partly on Guerrero's misinterpreting notions such as *acting from ignorance* and *culpable ignorance*. When these notions are properly understood (as I will suggest) the Moral Ignorance Thesis survives Guerrero's attack. According to Guerrero, something is done from ignorance if and only if¹⁵⁸ the following conditions are met:

(AI1) a person *S* performed action *A* while lacking belief in some proposition *p*, (AI2) *S* would not have done *A* had *S* believed that *p*, (AI3) *p* is true.¹⁵⁹

I cannot think of any reason to reject any of these three conditions, but in 2.2 I will suggest that they are not sufficient to act from ignorance.

Then Guerrero distinguishes three ways in which a person can be ignorant of

155 Guerrero (2007), p. 60.

156 Zimmerman (1997), Rosen (2004).

157 King talks about "epistemic condition" of responsibility in general. But his examples are about factual ignorance.

158 Guerrero does not use the biconditional but speaks of "natural understanding of" something done from ignorance.

159 Guerrero (2007), p. 62.

some fact F:

(I1) cases in which a person is ignorant because she has never thought about the issue (...);
(I2) cases in which a person is ignorant because, though she has thought about the issue, she has come to have false beliefs about F; (I3) cases in which a person is ignorant because, though she has thought about the issue, she doesn't know what to believe.¹⁶⁰

While I take this distinction to be plausible, I dispute Guerrero's view of culpable ignorance. The view goes as follows:

A person S is culpably ignorant with regard to some fact F just in case S is ignorant of F and **(CI1)** S has done something (...) or **(CI2)** S has failed to do something (...), and **(CI3)** S ought to have behaved differently with regard to (CI1) and/or (CI2), and **(CI4)** S's failure to behave differently with regard to (CI1) or (CI2) produced S's ignorance of F.¹⁶¹

I disagree with this view of culpable ignorance because I believe another condition should be added: **(CI5)** S thought that he ought to behave differently with regard to (CI1) and/or (CI2). The reason why this addition is necessary will hopefully become clear if we consider the following example. Suppose that I have to decide how to vote at the next general elections. I call my dad and ask him who he thinks is the best candidate. I then vote according to what my dad said. Suppose that calling my dad is the only action I think I should perform in order to decide how to vote. What I should in fact do, in a non-perspectival sense, is to collect more

160

Ibid., p. 63.

161

Ibid., p. 66.

information from different sources. Suppose that I end up voting for the worse candidate. It seems to me that I am not blameworthy for *not* knowing who the best candidate is, because at the time when I had to decide how to vote, I did not think that I should do more than just asking for my dad's opinion to find out who the best candidate is. I genuinely thought that asking for my dad's opinion would be sufficient. But I was unlucky: the candidate my dad suggested was the wrong one.

According to Guerrero's (CI3), since I should have collected more information regarding the presidential candidates, and given that (CI2) and (CI4) are satisfied, then I am culpably ignorant for not knowing who I should vote for at the election. But this seems mistaken. Whether I am culpably ignorant depends on whether I am culpable for thinking that asking for my dad's opinion is sufficient to decide how to vote. In other words, it depends on whether I performed an act that I knew would (or might) lead me to believe that asking for my dad's opinion is sufficient to establish how to vote (although I thought I should not perform that act). Suppose there is a pill that will make me believe that asking for my dad's opinion is sufficient to establish how to vote. I might take the pill to avoid feeling guilty for my behaviour. I cannot be bothered to do more research, so I need a way to be content with my action, although I know it is (or might be) wrong. In this case I am blameworthy for believing that asking for my dad's opinion is sufficient to establish how to vote. However, this looks like a highly hypothetical case, and I am inclined to think that in most cases I would be blameless for thinking that asking for my dad's opinion is sufficient to find out who the best presidential candidate is.

Guerrero endorses moral epistemic contextualism. This is the view that how

much we are morally required to do, from an epistemic point of view, with regard to investigating a certain proposition, varies depending on the moral context. As the relevant context becomes more morally serious, our epistemic obligations grow. Guerrero's idea is that we are blameworthy if we do not fulfil the epistemic obligations appropriate to the moral context.¹⁶²

Now since the Moral Ignorance Thesis is about culpability and not simply about obligatoriness, it seems to me that whether one's ignorance is blameworthy does not depend only on one's epistemic obligations, but also on whether the agent thought he ought to fulfil them in order to avoid that ignorance. If the agent decided not to fulfil the epistemic obligations he thought he should fulfil, and these turn out to be, in fact, the epistemic obligations he should fulfil, then the agent is blameworthy for his ignorance. Since he did not do what he thought he should do, which is also what he in fact should have done (in a non-perspectival sense) then he deserves blame for failing to fulfil his epistemic obligations, and for the resulting ignorance.

Suppose that I think I should do something else beyond asking for my dad's opinion on how to vote. And suppose that I think I should inform myself more because I believe that, otherwise, I might end up voting for the wrong candidate. In this case I am culpable for not knowing who the best presidential candidate is. Surely even if I informed myself properly, I could still get the candidate wrong. But in sticking only to what my dad suggests, I knowingly perform an action that I believe to be wrong. And if I perform it, there is a higher chance that I will end up holding a false belief (*i.e.* believing that "x is the best candidate", when x is not) and

performing a wrong action (*i.e.* voting for the wrong candidate). And that is why blame for my resulting ignorance is deserved. My claim is then that if we add (CI5) to Guerrero's definition of culpable ignorance, the Moral Ignorance Thesis holds true: whenever an agent acts from moral ignorance, she is culpable for the act only if she is culpable for the ignorance from which she acts.

2.1

In his paper "Culpability and Ignorance", Gideon Rosen discusses the case of an ancient slaveholder to show that there are cases of blameless moral ignorance.

(...) consider an ordinary Hittite lord. He buys and sells human beings, forces labour without compensation, and separates families to suit his purposes. (...) what he does is wrong. The landlord (...) thinks he is [entitled to do these things]. (...) if he had thought otherwise, he would have acted differently. (...) he acts from moral ignorance (...). (...) [In the context of] ancient Near Eastern slavery (...) one became a slave through bad luck or imprudence; in principle the status could befall almost anyone. (...) The institution of chattel slavery [and its legitimacy] was *simply taken for granted*.¹⁶³

Some support for the Moral Ignorance Thesis seems to stem from cases such as this, since the Slaveholder seems blameworthy for holding slaves only if his ignorance were culpable. The slaveholder is not blameworthy for holding slaves and he is blamelessly ignorant of the wrongness of slavery. And he is blameless for

holding slaves exactly *because* he is blamelessly ignorant of the wrongness of slavery. Guerrero intends to undermine the support for the Moral Ignorance Thesis that stems from cases such as this. He wants to suggest that this example does not support the Moral Ignorance Thesis, since it is more reasonable to see the ancient slaveholder as blameworthy. Guerrero denies both that the slaveholder is blamelessly ignorant and that he is blameless for holding slaves.

Suppose that the ancient slaveholder has thought about whether slavery is morally permissible and he has come to believe that it is indeed permissible.¹⁶⁴ Guerrero notes that since the moral significance of the actions involved in slavery is high, the epistemic obligations regarding the investigation of whether slavery is morally permissible are also high. And this makes it implausible that the slaveholder has met them. Therefore, Guerrero claims that it is reasonable to see the slaveholder as blameworthy. In addition, since determining that slavery is wrong does not seem difficult in terms of reasoning, it is natural to see the slaveholder as culpably ignorant.¹⁶⁵

However, it seems to me that even if we assume that it is indeed easy to rationally conclude that slavery is wrong, the slaveholder could still have made a mistake in his reasoning, he could be blameless for this mistake, and blameless for reaching the wrong conclusion about slavery. If this is so, the slaveholder may indeed be blamelessly ignorant, and the case will rightly suggest that the slaveholder would be culpable for holding slaves only if he were culpable for not knowing that slavery is wrong.

164

Guerrero (2007), p. 70.

165

Ibid., pp. 71-72.

Suppose that the slaveholder thinks that what he should do to establish whether slavery is permissible is to ask the wisest man in town. In doing this, he does not fulfil what are in fact his epistemic obligations, but he fulfils what he *genuinely takes to be* his epistemic obligations. Suppose the wise man suggests that slavery is permissible. If so, the slaveholder is blameless for not knowing that slavery is wrong, because he thought he was fulfilling his epistemic obligations. In fact, he didn't fulfil them. And in doing so he did something wrong. He performed a wrong but blameless action.

It appears to me that there is a substantial difference in terms of the evaluation of the agent between an agent who violates an obligation because of a mistake in her reasoning, and an agent who violates an obligation deliberately, acting against her better judgment. When I compare propositions (*e.g.* “slavery is permissible” and “slavery is impermissible”), I can make mistakes in my reasoning over which I have virtually no control (I can only broadly train myself to perfect my reasoning). But when I make a choice I seem to have full control over what to do or refrain from doing (*e.g.* informing myself). As I inform myself, if I make a mistaken inference, I am, in normal cases, not culpable for that mistake. But when I choose not to inform myself (although I think I should) because I cannot be bothered, it is clear that I deserve some blame. And the reason is that it seems that I exercise more (direct) control over my act of choice than over my reasoning.

In order to solve an equation it is not sufficient that I simply *will* to solve it. I do not have full control over whether I will get the right result. I can only study

maths and practice on several similar types of equations. But if I want to sit down, study maths and think about how to solve the equation, all I need to do is simply to choose to do it, as it seems that I have greater control over these decisions (some external force may prevent me from performing the action, but not from making the decision, at least in normal cases). If I think I should do my homework (which includes solving equations) but I choose not to approach the equations, I may deserve some blame for not trying (if that is what I in fact should do, in a non-perspectival sense). But if I try, I won't deserve blame for making mistakes. I might deserve blame for not practising enough, but that blame will not attach to a specific mistake I make. For instance, I would deserve blame in a case in which I did not study a certain formula (that I thought I ought to study) and I foresaw that I might fail to apply it in the equation, if I don't memorise it. The same applies to moral ignorance: I have control over deciding to inform myself about a moral matter and to act according to the belief I come to hold. But I do not have the same kind of control (which is relevant to the attribution of blameworthiness, I argue) over making a mistake in my reasoning or in my collecting evidence (if done genuinely), and I don't deserve blame for ending up having a wrong moral belief as a result of that mistake.

Then Guerrero considers a scenario in which the slaveholder has not thought about whether slavery is morally permissible. Again, Guerrero claims that given the moral seriousness of the actions involved in slavery, it is hard to imagine that the slaveholder has never thought about whether what he does is impermissible. In addition, Guerrero asserts that even if it were the case that the slaveholder has never thought about the issue, this would give reason to see him as culpable rather than the opposite.¹⁶⁶

However, it seems to me that the fact that the slaveholder fails to think about whether slavery is permissible does not count, per se, as a reason to see him as culpable, because failing to think about the permissibility of slavery can be itself culpable or non-culpable. Suppose I am planning to watch a film later tonight. I entertain the idea of streaming it illegally. Suppose that there is a pill that will make me forget to think about whether what I consider doing is wrong. I take the pill so that I will not go through the process of considering whether streaming films illegally is permissible, so that I will not feel guilty for streaming the film illegally if I conclude that it is wrong to do so. I go ahead and stream the film. In this case I am blameworthy for my ignorance, because I have decided, against my better judgement, to perform an action that will prevent me from thinking about the moral permissibility of illegal streaming. But if the case of the slaveholder is not a case such as this, the slaveholder does not seem blameworthy for not having thought about the moral permissibility of slavery. And if this is the case then Guerrero does not seem to have successfully shown that cases such as the case of the slaveholder do not support the Moral Ignorance Thesis.

2.2

Finally, Guerrero presents a case that is intended to be a counterexample to the Moral Ignorance Thesis:

Douglas is contemplating killing a pig for dinner (imagine that the pig is currently still alive

and healthy). (...) he is able to understand the idea of moral status, and is questioning whether or not the pig has significant moral status – moral status such that it would be wrong to kill the pig in order to eat it. If he knew the pig did have significant moral status of this sort, Douglas would not kill it. After some effort, he is still unable to come to an answer to this question – he doesn't know whether the pig has significant moral status. His reasons for wanting to kill and eat the pig are straightforward: he likes the taste. There are plenty of other food options that he knows are morally permissible, which would be nutritious and which he would enjoy eating, and which would not require killing pigs or any other animals. Still unable to decide the question, he goes ahead and decides to kill and eat the pig, and does so successfully. (...) let us stipulate that pigs in fact have significant moral status, such that it is wrong to kill them in order to eat them.¹⁶⁷

Douglas doesn't know whether the pig has significant moral status. He is aware of his ignorance and he knows that it would be wrong to kill the pig if it turned out to have significant moral status. Moreover, he knows he does not have compelling reasons – which *would* outweigh the pig's having moral status – to eat the pig.¹⁶⁸ Guerrero claims that this is a case in which the agent acts *from* ignorance, and he is *not* culpable for that ignorance but he *is* culpable for his action. Therefore, according to Guerrero, this is a counterexample to the Moral Ignorance Thesis.

I disagree with Guerrero as it seems to me that the example of Douglas and the pig is not a counterexample to the Moral Ignorance Thesis. When Douglas acts he does not know whether the pig has moral status. He is also not culpable for not knowing that the pig has moral status. But Douglas is nonetheless culpable for killing

167

Ibid., p. 76.

168

Ibid., pp. 77-78. Let's stipulate that it is in fact true that liking the taste of the meat, as a reason to kill the pig, is outweighed by the fact that pigs have significant moral status.

the pig because he does not act *from ignorance of that proposition*. He acts *knowing* that it might turn out that the pig has moral status. He is then aware that he is performing a risky action and that he may end up committing a wrong action. Therefore, since Douglas acts in ignorance of the proposition that pigs have moral status but not *from* ignorance of that proposition, the case is not a counterexample to the Moral Ignorance Thesis. It is not a counterexample to the Moral Ignorance Thesis because this is not a case of ignorant action, strictly speaking.

My disagreement with Guerrero might derive from a disagreement about how to understand the notion of ‘acting from ignorance’. Let us remind ourselves of Guerrero’s formulation: The agent acts from ignorance when (AI1) she performs the action while lacking belief in some proposition, (AI2) she would not have performed that action had she held that belief, and (AI3) the proposition is true. I shall suggest that these conditions do not allow us to correctly assess Douglas’ case. Douglas kills the pig while lacking knowledge of the proposition “pigs have moral status” (AI1). He would not have killed had he believed that pigs have moral status (AI2). Also, we stipulated that the proposition “pigs have moral status” is true (AI3). This seems then to be a case that meets Guerrero’s conditions for something to be done from ignorance. However, note that Douglas knows something else that leads him to conclude that he should not kill the pig. He knows that there are no compelling reasons to take that risk. So he knows he should not kill the pig *not* because pigs have moral status, but because pigs *might* have moral status, and he has no compelling reasons to take the risk of ending up killing a being with moral status. Therefore, the fact that Douglas lacks the belief that pigs have moral status does not

play any actual role in his action. What he knows is enough for him to be able to conclude that he should not kill the pig.

Consider a different version of the case. Suppose that Douglas must decide whether to kill a pig in order to eat it. He believes that pigs are smarter than humans. Let's stipulate that this is in fact *not* the case. Suppose that if Douglas knew that pigs are less smart than humans, he would not kill the pig. But Douglas knows something else: he knows that pigs have moral status. And this is enough to make him conclude that he should not kill the pig. Douglas goes ahead and kills the pig. It seems to me that in this version of the case we don't want to say that Douglas acts from ignorance. He rather acts against his better judgment. He acts knowing that what he should do is not to kill the pig. And it seems to me that the same applies to the original case. It is true that Douglas does not know about the moral status of the pig. But this doesn't really play a role in his action, because he also holds the belief that, since he does not have compelling reason to risk killing something with moral status, he should not kill the pig. And this is why he is blameworthy for killing the pig.

It seems then that in order to accurately evaluate cases such as Douglas, Guerrero's notion of 'something done from ignorance' could be revised as follows: something is done from ignorance if and only if (AI1) a person *S* performed action *A* while lacking belief in some proposition *p*; (AI2) *S* would not have done *A* had *S* believed that *p*; (AI3) *p* is true; and **(AI4) *S*'s lacking the belief in *p* played a role in doing *A*.** Guerrero might have thought that (AI4) is already entailed by (AI2), but I have shown that this is not cases. I have shown that the counterfactual in (AI2) might be true, and yet it be the case that the lack of belief played no role in bringing

about the action.

In her paper “Does Moral Ignorance Exculpate?” Elizabeth Harman considers a similar case of uncertainty: “Alice spoons cyanide into Bob’s coffee. Alice is *ignorant* that she is spooning cyanide, in that she lacks a belief that it is cyanide; rather, she has .5 credence that it is sugar and .5 credence that it is cyanide”.¹⁶⁹ Harman claims that Alice is blameworthy for spooning cyanide even though she is not blameworthy for her ignorance. And similarly, Harman suggests, when someone acts wrongly while unsure whether what she is doing is wrong, she may be blameworthy simply for performing that action, and not because she is blameworthy for being unsure.

This case looks similar to Guerrero’s case of the pig. Alice does not know whether she is spooning sugar or cyanide, but she probably knows something else (the example is under-described), which is crucial in the moral evaluation of the case. She may know that there are no compelling reasons for taking the risk of spooning cyanide into Bob’s coffee (it would be worthy to take the risk if Bob were about to die and, instead of sugar, Alice could spoon a powerful antidote that would save Bob’s life). So when Alice spoons cyanide she may not act from ignorance. On the contrary, she may act knowing that she is taking a risk that should not be taken, because of the absence of compelling reasons. If Alice blamelessly lacked the belief that the risk of spooning cyanide should not be taken, then she would be blameless for spooning cyanide into Bob’s coffee. Therefore, Alice’s case, like Douglas’, does not support the thesis that one can be blameworthy for an action performed from ignorance even if they are not blameworthy for their ignorance.

3.0

In conclusion, in this chapter I have undermined some objections to the thesis that whenever an agent acts from ignorance, she is culpable for the act only if she is culpable for the ignorance from which she acts. My defence applies both to cases of factual ignorance and to cases of moral ignorance.

In the first part of the chapter (1.0-1.2) I have focused on factual ignorance. I have undermined King's argument for the claim that tracing is not necessary to explain responsibility for ignorant behaviour by showing that, if we carefully examine the examples presented by King, we realise that it is not at all clear that they support his claim. The reason for this is that even in these cases the agent may indeed be seen as responsible for his ignorant behaviour because he is responsible for some previous act that caused that ignorance. I have emphasised that, looking at King's examples, an issue of individuation of the action may arise, and this makes the cases more ambiguous and hard to assess, especially for actions particularly extended in time. I have also criticised King's evaluation of a case of ignorant behaviour performed because of negligence, suggesting that even in cases of negligence the agent may be blameless for her ignorance.

In the second part of the chapter (2.0-2.2) I have focused on moral ignorance. I have undermined Guerrero's argument against the Moral Ignorance Thesis, which claims that whenever an agent acts from moral ignorance she is culpable for the act only if she is culpable for the ignorance from which she acts. I have claimed that

Guerrero's attack on the Moral Ignorance Thesis depends partly on Guerrero's misinterpreting notions such as *acting from ignorance* and *culpable ignorance*, and that when these notions are properly understood the Moral Ignorance Thesis survives Guerrero's attack. Moreover, I have criticised Guerrero's claim that cases such as the Ancient Slaveholder do not provide support to the Moral Ignorance Thesis. Finally, I have shown that Guerrero fails to provide a strong counterexample to the Moral Ignorance Thesis. I have suggested that the case he discusses fails to support his claim, since it is *not* a case in which the agent acts from ignorance, while he is culpable for his action, but not for his ignorance. Actually, in Guerrero's case, the agent does *not* act from ignorance. Guerrero disagrees with this claim because his notion of acting from ignorance is mistaken. When we fix it, it becomes clear that the case discussed by Guerrero is not a counterexample to the Moral Ignorance Thesis.

Chapter 5 – Culpability and Akrasia

In this chapter I will undermine some objections to the thesis that moral culpability always involves akrasia. I will call this the *Akrasia Thesis*. An akratic action occurs when I act against my better judgement. Suppose I must decide whether to do A or B. I believe I should do A, but I do B regardless of my belief. This is an akratic action. Most philosophers believe that sometimes agents are culpable for their actions. In the literature we find two main views on the connection between culpability and ignorance, with

respect to akrasia. One view is that culpability always involves akrasia, the view I will defend.¹⁷⁰ We are culpable only if we acted akratically or if we are culpably ignorant for our action, where culpable ignorance is ignorance that derives from an earlier akratic action. The alternative view is that we can be culpable for an action even if it is not akratic or resulting from an earlier akratic action.¹⁷¹

In 1.0 I will present Zimmerman's view on culpability and ignorance, which implies that the Akrasia Thesis is true. In 1.1 and 1.2 I will undermine the objections to Zimmerman's view raised by Montmarquet and Peels, respectively. Finally, in 2.0, I will examine George Sher's considerations on moral responsibility which might seem to cause problems for views such as the Akrasia Thesis. I will argue that Sher's considerations do not undermine the Akrasia Thesis as his criticisms only target views that call for a full identification of the judge with the agent. I will claim that my view is not as problematic as the views of responsibility to which Sher objects.

1.0

In this section I will explain Michael Zimmerman's view on culpability and ignorance. According to this view, since we have only indirect control over our ignorance, culpability for ignorance is always indirect. In addition, all culpability is traced to culpability involving the agent's occurrent belief that she is doing something wrong.

In his paper "Moral Responsibility and Ignorance", Zimmerman examines the following case:

170

Zimmerman (1997), Rosen (2004), Levy (2009).

171

Montmarquet (1999), FitzPatrick (2008), Sher (2009).

Doris has just driven her car into a tree. She's unconscious, slumped over the steering wheel. Perry comes upon the scene. He's shocked at what he finds. He looks around to see if anyone can help, but there's no one else there. He's wondering whether he should (...) call for aid, but visions of wrecked cars catching fire and exploding into roiling balls of flame fill his mind, and he feels that he must rescue the driver now or else she'll surely die. So, with considerable trepidation, Perry rushes in and quickly drags Doris free from the wreck, thinking that at any moment both he and she might get caught in the explosion. (...) As it happens, the car does not explode. (...) Soon after, some emergency vehicles screech to a halt. (...) the paramedics take a look at Doris, and they arrive at a chilling conclusion: Perry has paralyzed Doris.¹⁷²

Suppose that Perry's dragging Doris free from the wreck was morally wrong, in a non-perspectival sense. Perry was ignorant of the wrongness of his action because he was ignorant of the risk of paralysis. Zimmerman gives the following argument:

- (1) Perry is culpable for behaving as he did (...). [supposition]
- (2) One is culpable for behaving ignorantly only if one is culpable for being ignorant.¹⁷³
- (3) Perry is culpable for being ignorant.
- (4) One is culpable for something only if one was in control of that thing (...).
- (5) Perry was in control of being ignorant.
- (6) No one can be in direct control of being ignorant.
- (7) If one is in control of something but not directly so, then one is in indirect (...) control of it.
- (8) Perry was in indirect (...) control of being ignorant.
- (9) If one is culpable for something but was only in indirect control of it, then one is indirectly

172 Zimmerman (1997), p. 410.

173 In chapter 4 I have undermined objections to this claim.

(...) culpable for it.

(10) Perry is indirectly (...) culpable for being ignorant.

(11) If one is indirectly (...) culpable for something X, then one is directly culpable for something else, Y, of which X was a consequence.

(12) Perry is directly culpable for something of which his ignorance was a consequence.¹⁷⁴

Zimmerman then raises the question of what could be that something of which Perry's ignorance was a consequence. It cannot be his carelessness or inconsiderateness – Zimmerman claims – because Perry is not aware that he is being careless or inconsiderate in acting wrongly. Zimmerman is not implicitly claiming that an agent can never be directly culpable for his carelessness. Zimmerman believes instead that an agent can be blameworthy for his carelessness only if the agent was aware that he was being careless. Montmarquet, on the other hand, thinks that Perry may well be directly culpable for his carelessness. And not just for careless acting, but for carelessly *forming a false belief*, regardless of whether he is aware of his carelessness.¹⁷⁵ Zimmerman claims that Perry cannot be directly culpable for his inattentiveness either, because he is ignorant of the need to pay attention not to paralyse Doris and he does not know that it is wrong not to pay such attention. Zimmerman then concludes that there is nothing for which Perry is directly culpable and whose wrongness he is ignorant of.¹⁷⁶ Perry could be culpable for this thing only if he were culpable for the state of ignorance he is in. And he could only be indirectly culpable, not directly culpable, for this ignorance.

Zimmerman's thesis is that we can be culpable only for what is under our control and since we are never in direct control of our ignorance, culpability for ignorance is *always*

174 Zimmerman (1997), pp. 414-415.

175 I will discuss this in section 1.1.

176 Zimmerman (1997), pp. 416-417.

indirect.¹⁷⁷ Being indirectly culpable for something means being directly culpable for something else. And this cannot be ignorant behaviour. Given that Zimmerman seems to assume that an agent can only do wrong either ignorantly or akratically (and that there is no third option), all culpability is then traced to culpability involving the agent's belief that he is doing something wrong.¹⁷⁸

As an additional claim, Zimmerman states that for Perry to be culpable for his ignorant behaviour there must also be some *cognitive connection* between his ignorance and another state or event for which he is directly culpable and of which his ignorance is a consequence. In order to clarify what Zimmerman means by 'cognitive connection', let us think of a case in which this cognitive connection principle is clearly not satisfied. Suppose I am in a bar waiting for my best friend. I have promised I would hang out with her as she needs some advice from me about a difficult choice she must make. I start chatting with a stranger but the music is quite loud and I soon develop a strong headache. I ask the guy I have just met if he happens to have something against headache. What he offers me is a pill which contains animal-derived ingredients. I am vegan and I think it would be morally wrong to take the pill. Nonetheless, I accept the offer. Unbeknownst to me, the pill has the effect of making me forget about my promise to wait for my best friend and hang out with her. A few minutes after taking the pill, I leave the bar and go back home. This is a case in which my failing to keep my promise is not *cognitively connected* to my taking the pill, because the reason I take the pill is to get rid of my headache. My action causes my state of

177 Zimmerman does not specify the conditions for direct control. I take direct control to be, roughly, about actions and indirect control about states or events. I can deliberately collect evidence poorly. That is something over which I have direct control. That will make me ignorant of certain important pieces of evidence. That ignorance results from my action of collecting evidence poorly and it is a state over which I have only indirect control. In general, my control over my forming a certain belief is always less direct than the one I have over my actions.

178 Zimmerman (1997), pp. 417-418.

ignorance, but my ignorance is not cognitively connected with my action, so there is no transfer of responsibility from taking the pill to forgetting my promise. I am not responsible for that state of ignorance.

Back to Perry's case, Zimmerman supposes that there is some cognitive connection between Perry's ignorance and some previous action of his:

Imagine (...) that Perry had signed up for some training in first aid but had deliberately decided to skip the session having to do with the treatment of accident victims, despite believing that he ought not to skip it precisely because he would remain ignorant about how to treat such victims.¹⁷⁹

Zimmerman claims that in this case Perry may well be culpable for skipping the session and indirectly culpable for the resulting ignorance.¹⁸⁰ According to Zimmerman, this cognitive connection must involve occurrent beliefs about wrongdoing. When Zimmerman speaks of occurrent beliefs, he means beliefs we hold consciously before our mind. These are the kind of beliefs that are required for culpability. As opposed to occurrent beliefs, beliefs that are merely dispositional cannot be acted upon with the intention to heed or not to heed them. One cannot act deliberately on or deliberately despite a dispositional belief. One is culpable for one's action only if their beliefs about wrongdoing play a role in the reason why one performs the action. And only occurrent beliefs play such a role, according to Zimmerman.¹⁸¹

179

Ibid., p. 420.

180

Ibid., pp. 420-421.

181

Ibid., pp. 421-422. This claim is rejected by Rik Peels who argues that also unactivated dispositional beliefs may play a role in explaining culpability, as I will discuss in 1.2.

To clarify Zimmerman's condition of cognitive connection further, it is helpful to note that according to Zimmerman when an agent performs x at t_1 and he foresees that what he is doing will cause a state of ignorance at t_2 , he will be responsible for that state of ignorance. If he thinks that that ignorance is a mental state in which he should not be (because it might lead to performing wrong actions at t_3), then the agent is blameworthy for that ignorance. If he knows he might perform a wrong action at t_3 as a consequence of his ignorance, he may be blameworthy for performing a wrong action at t_3 (supposing that he knows that one action between y and z is wrong at t_3 and that his state of ignorance will prevent him from seeing which is the action to perform).

1.1

In this section I will discuss the objections to Zimmerman's view raised by James Montmarquet. In his paper "Zimmerman on Culpable Ignorance", Montmarquet does not explicitly dispute any of the premises (1)-(11) of Zimmerman's argument. However, he criticises Zimmerman's view. Zimmerman claims that Perry cannot be directly culpable for his carelessness because he is not aware that he is acting carelessly in rescuing Doris. Montmarquet claims instead that Perry can be directly culpable not only for his carelessness in acting (even if he is not aware that he is acting carelessly), but most importantly for his carelessness in forming the belief that

he should move Doris in the first place. Note that this does not amount to denying (6), because Montmarquet believes that we may be directly culpable for carelessly forming a belief but still only indirectly culpable for the state of ignorance that follows.

In order to show that Perry may well be directly culpable for carelessly forming the belief that he should move Doris, Montmarquet draws a parallel between believing carelessly and ‘running quickly’. He claims that a person is in direct control both over the speed with which she runs and over the running itself.¹⁸² Montmarquet claims that Zimmerman’s argument seems to suggest a reading of this case according to which a person would have only indirect control over her running, as a consequence of being in direct control of her speed. And this appears to be wrong because it is not the case that we cause our running by speeding up. And this remains true even if we describe one’s running quickly as a consequence of the effort one is exerting to run that quickly. Montmarquet concludes by suggesting that also in the case of forming a careless belief, the believer has direct control both over her degree of care and over her forming the belief.¹⁸³ Running quickly is like believing carelessly, in the sense that an individual can have direct control over her degree of care and over her forming a belief in the same way she has direct control over her running and over the speed with which she runs. Montmarquet would then claim, contrary to Zimmerman, that Perry’s carelessness in forming the belief that he should move Doris out of the wreck could indeed be that something over which Perry had direct control (and of which Perry’s ignorance was a consequence), regardless of the

182 I take it that Montmarquet means that one can decide to increase or decrease the speed with which one runs as directly as one can decide to start or stop running.

183 Montmarquet (1999), pp. 843-844.

fact that Perry was not aware that he was being careless.

However, I think that Montmarquet's parallel does not hold up. Take the case of running. I might think I should run because otherwise I will miss the bus. Or I might think I should run faster because I see the bus coming. In both cases it sounds reasonable to say that I have direct control over running and over running faster. However, Montmarquet fails to see that the case of forming a belief is different. In order to come to believe, for example, that I am late for the bus, I most likely need to check the time. I express some degree of care in checking the time and in making sure that I am not late for the bus. But my thinking that I should check the time most likely depends on whether I think that I might be late for the bus. And my thinking that I might be late for the bus may result from some other belief, *e.g.* that I have been late before. And so on. I have only indirect control over believing that I should leave now to avoid missing the bus. But I have direct control over checking the time. I might have most (internal) reason to check the time but still not do it if I cannot be bothered. I can still decide whether to do it or not. I might have most reason to believe that I might miss the bus if I don't leave now, but I cannot decide to (carefully) form the belief that I will miss the bus if I don't leave now.

For this view I draw from Bernard Williams. In his paper "Deciding to Believe" Williams claims that believing something implies that that something can be assessed as true or false. In other words, believing that *p* means to believe that *p* is true. When I say that I believe that *p*, I am claiming that *p* is true.¹⁸⁴ And it is because of this that we cannot decide to believe. If we could, we would be able to do so regardless of whether we believe *p* to be true, and we would know that we could

form a false belief. But if I can decide to believe, it is not clear that before forming the belief I would think of it as such, that is as something that represents reality. I could not know of a belief that I had decided to form it, because in order to decide to form a belief I must believe that I can do it.¹⁸⁵ Of course Williams is aware that there are ways I can bring about the formation of a certain belief, such as through hypnotism or by taking drugs. However, Williams claims there are two senses in which one may want to believe something: in terms of ‘truth-centred motives’ or ‘non-truth centred motives’;¹⁸⁶

Suppose a man’s son has apparently been killed in an accident. It is not absolutely certain he has, but there is very strong evidence that his son was drowned at sea. This man very much wants to believe that his son is alive.¹⁸⁷

Williams notes that what the man actually wants here is for his son to be alive. In other words, he wants the belief that his son is alive to be true. Williams calls this a truth-centred motive. Turning to a hypnotist, so that he can be made to believe that his son is alive, will not get him what he wants.

(...) if he said, ‘Well, of course what I would like best of all is for my son to be alive; but I cannot change the world in this respect. The point is, though that even if my son isn’t alive, I want, I need to believe that he is, because I am so intolerably miserable knowing that he isn’t’.¹⁸⁸

185 *Ibid.*, pp. 147-148.

186 *Ibid.*, pp. 148-149.

187 *Ibid.*, pp. 149-150.

188 *Ibid.*, p. 150.

According to Williams' distinction, this is a non-truth-centred motive. In this case it would make sense for the man to use drugs or hypnotism to bring about the belief that his son is alive. The man will be in direct control of, for example, taking the drug, and in indirect control of acquiring the belief that his son is alive. Thus, forming a belief is something over which I have only indirect control, or at least a kind of control that is less direct than the control I have over actions such as taking a drug or, back to Montmarquet's example, running quickly.

Referring back to Perry's case, I agree with Montmarquet that the fact that Perry's belief is a consequence of his lack of due care is not the reason why he has only indirect control over forming that belief. Running is indeed the consequence of some physical mechanism, but in normal cases I can run if I decide to, whereas I cannot simply decide to form a belief. I can decide to inform myself and then the belief I form will depend on what evidence I have acquired and on my understanding of it. Forming a belief is then a more passive action compared to running, as what is active is not really forming the belief but rather collecting information.

When Perry decides whether to sign up to the first-aid training, he knows that he would learn what to do in cases of car accidents. But he does not know what belief he will form about what to do in cases of car accidents. That is something he will find out only by attending the session. Suppose the first-aid training were something morally required for Perry to take and that he knows that. He has direct control over informing himself by going to the session, but he has only indirect control over forming the belief that in the case of an accident like Doris' there is a risk of paralysis if you drag the person free from the

wreck. Thus, if Perry does not attend the training in this case, he will be blameworthy for not attending it because he has direct control over doing so. And he will be blameworthy for being ignorant about how to assist someone in the case of an accident because he is indirectly in control of acquiring the belief that is needed in cases such as Doris' accident. It is true that Perry might still attend the session and not understand what to do in situations such as Doris', but by not attending he prevents himself from being exposed to the knowledge he needs in order to be able to perform the right action (when he finds himself in the relevant moral situation). And that is the main reason why he is culpable for his ignorance. Note that it is not the lack of predictability that explains the passivity of forming beliefs. What makes the process of forming the belief (that in cases of accident Perry should not move the victim) passive is not that Perry could not predict it before attending the course. Lack of predictability is simply a further feature of the first-aid course scenario.

Montmarquet claims that Perry may be *directly* culpable for forming the careless belief that he should move Doris away from the car. This belief is careless because Perry should have asked himself whether it is wrong to move Doris away from the car. Montmarquet rejects then Zimmerman's claim that Perry's carelessness is not that something for which Perry is directly culpable and of which Perry's ignorance is a consequence (recall that what Perry ignores is that he should not rescue Doris because of the risk of paralysis). Montmarquet claims that Perry can (and is expected to) be in a different mental state and his being in a wrongful mental state without being aware that that state is wrongful can be still directly culpable. This is due to the fact that he might not have bothered to ask himself whether that mental state is wrongful. And that is culpable.¹⁸⁹

Although I agree that one's lack of awareness that something is wrong can be

culpable, Montmarquet seems to overlook that not asking oneself whether something is wrong may be blameless. It is true that Perry should ask himself whether what he is doing is wrong, but maybe only in the sense that it is morally desirable that he asks himself that.¹⁹⁰ It is morally desirable because if he asks himself whether what he is doing is wrong he might conclude that it is better not to move Doris, and he might consequently refrain from performing a morally wrong action. Perry *should* know that moving Doris might paralyse her in the sense that it is morally desirable that he knows that, so that he will not harm Doris. And naturally we do not want Doris to get paralysed. However, Perry cannot be expected to know that moving Doris is wrong if he has not asked himself whether it is wrong in the first place, and it might not be reasonable to expect him to ask himself whether it is wrong to move Doris. For instance, he might have never experienced that kind of situation. If his ignorance does not result from an earlier akratic act, an act he chose to perform (or not to perform) knowing he will become ignorant, then he will not be culpable for not asking himself whether it is wrong to move Doris.

There is then an ambiguity in the sense of ‘should’ employed when we evaluate this kind of cases. Zimmerman could concede that Perry should have asked himself whether what he was doing was wrong, but only in the desirability sense of ‘should’. Zimmerman cannot concede that if Perry should have asked himself whether what he was doing was wrong then it must be the case that he is culpable for it. I acknowledge that there may be a different sense of ‘should’ which does not simply mean that it is morally desirable to perform an action. But once we disambiguate different senses of ‘should’, Montmarquet’s objection seems less strong, since the reason why people will see Perry as culpable may be

190 Note that there are cases of a ‘desirability’ sense of ‘should’ where the desirability is not moral desirability (e.g. if I say that I should have done *x* because *x* would have had the better outcome for me).

because they confuse different senses of ‘should’ and generally tend to conflate cases that should be separated as far as the culpability of the agent is concerned.

Zimmerman’s suggestion is that if we think harder about Perry’s case we see that he may not be culpable for not asking himself whether it is wrong to move Doris. After all, Perry’s not bothering to ask himself whether believing that he should rescue Doris was a wrongful state of mind can be interpreted as follows: Perry thought that he should rescue Doris but he also thought that he should ask himself whether his state of mind is wrongful, but he decided to go ahead anyway and rescue Doris. Either Perry thought that going ahead and rescuing Doris was wrong or he thought that going ahead and rescuing Doris was right. If he thought it was wrong, he is culpable for rescuing Doris, exactly because believing this means believing that he should not rescue Doris. If he thought that going ahead and rescuing Doris was the right thing to do rather than stopping to ask whether that belief is false (*e.g.* because he thinks he is running out of time and the car could explode any minute), he is blameless for rescuing Doris and for paralysing her, because he acted according to his better judgement.

Montmarquet disagrees with Zimmerman’s claim that there is nothing for which Perry is directly culpable and about the wrongness of which he is ignorant. The reason for this disagreement is that Montmarquet believes that Perry might have been directly culpable for not being “open to truth-and value-related considerations”, which is a fundamental expectation, Montmarquet claims. One may then be blameworthy for not being open to truth-and value-related considerations and for not being open to the need to be open.¹⁹¹ And this is something over which we have, to some extent, direct control, in the

191 I assume that the idea of being "open to the need to be open" is a higher level attitude equivalent to the idea of considering being sensitive to truth-and value-related considerations. I find it highly implausible that we have any kind of control over such a state.

sense that we can order ourselves to be open even if the conclusions we are led to are not under our control.¹⁹²

I disagree with Montmarquet's claim that we have direct control over our openness to truth-and-value-related considerations. And if we do not have such control, then we cannot be directly culpable for not being open to truth-and-value-related considerations. It seems to me that direct control is exercised only at the point when we decide what to do with our beliefs. We hold certain beliefs because we are open. We sometimes ask ourselves what we should do and realise what beliefs we have about what to do. If Perry is an agent who has been asking himself whether what he was doing was wrong on several occasions throughout his life, he is an agent who is open to truth-and value-related considerations. This remains true even if he failed to ask himself whether moving Doris is wrong on that particular occasion. He can still be considered a reflective moral agent who has simply failed to ask himself, on that particular occasion, whether moving Doris was something wrong to do. And he has done so blamelessly, since he has not deliberately brought about this failure. Moreover, contrary to what Montmarquet claims, thinking about needing to ask ourselves whether what we are doing is wrong can hardly be seen as something that we tell ourselves to do. If I am not open to the need to be open to truth-and value-related considerations, I might be a seriously defective moral agent, or I might not be a moral agent at all. But it is not clear why I would be blameworthy for it. Montmarquet might say that Perry has an obligation to be open to truth-and-value-related considerations in relevant occasions such as in the case of Doris' accident. But this does not seem to help Montmarquet's argument because that would be

equivalent to saying that Perry is culpable for not asking himself whether what he is doing is wrong in situations where he should ask himself whether what he is doing is wrong, or where a good moral agent would ask herself whether what she is doing is wrong (assuming that a good moral agent would be open to truth-and-value-related considerations in situations such as Doris'). But this kind of answer still overlooks the distinction between culpability and simply failing to fulfil the obligation of asking oneself whether what one is doing is wrong, a failure that may be non-culpable.

In conclusion, I think Montmarquet does not successfully reject Zimmerman's view that there is nothing for which an agent is directly culpable and about the wrongness of which he is ignorant. First, because Montmarquet's example of running quickly cannot appropriately be compared to that of believing carelessly. Secondly, because Montmarquet overlooks the fact that not bothering to ask oneself whether something is wrong can itself be blameless. And thirdly, because Montmarquet seems to ignore the possibility of an agent being open to truth-and value-related considerations and, at the same time, blameless for not asking herself whether what she is doing is wrong.

1.2

In this section I will consider another objection to Zimmerman's view. Some criticism of Zimmerman's view comes from Rik Peels' paper, "Tracing Culpable Ignorance". Zimmerman claims that for Perry to be culpable for his ignorant behaviour there must be

some *cognitive connection* between his ignorance and another state or event for which he is directly culpable and of which his ignorance is a consequence. This cognitive connection must involve occurrent beliefs about wrongdoing. And this is why culpability always involves akrasia. On the contrary, Peels claims that occurrent beliefs are not the only beliefs relevant to assessing culpability. One can be culpable for not activating dispositional beliefs about one's circumstances or about an action's normative status. For instance, if I forget to keep a promise, according to Peels I am blameworthy for not doing what I promised, in the absence of other excusing circumstances. This is because I have a *dispositional unactivated belief* that I should do what I have promised. According to Peels, I could and should have activated that belief. Failing to activate my dispositional belief that I should do what I have promised to do is neither an akratic omission nor an ignorant omission. Still, I appear to be blameworthy for failing to keep my promise, Peels claims. Therefore, culpability does not always involve akrasia, contrary to Zimmerman's claim.¹⁹³

However, it seems to me that cases of forgetting are cases that may exculpate exactly because of the absence of an occurrent belief, since I may blamelessly fail to activate a dispositional belief. The failure in belief activation may be blameworthy if I had an occurrent belief and performed an akratic action. Suppose I forget that I have promised to pick up my friend at the train station. I might recall that promise when my friend rings me saying he is waiting for me. At that point it would be appropriate for me to react by saying: 'Sorry! I should have remembered it!' But here there is an ambiguity both in how I use 'sorry' and in how I use 'should have'. When I say sorry I might simply be expressing sympathetic feelings towards my friend, for instance, sadness

about what has happened. By saying I am sorry I might simply acknowledge that I was causally responsible for what has happened and that I think it is a shame that that was the case. Alternatively, by saying I am sorry I might express my belief that I am culpable for what I have done and that I hope my friend will forgive me.

When I say that I should have remembered to pick up my friend, I might simply mean that not forgetting about it would have avoided the inconvenience I have now caused to my friend. And not causing any inconvenience to my friend is something morally desirable, in the sense that, from a moral point of view, we want a world in which I remember my promise to pick up my friend. Still, I happened to forget about that promise; that is, I did not activate that dispositional belief. Of course in a sense I could have avoided that situation. I could have remembered what I had to remember. I could have activated that belief. But suppose that when I made the promise I thought I would easily remember to keep that promise. I thought I would not need any reminder. I had no internal reason to think I would forget to pick up my friend. If this is how things stand, I think it is unreasonable to expect me to do something to avoid forgetting to pick up my friend. Therefore, it seems I am not blameworthy for forgetting to pick up my friend and for not picking up my friend. I concede that I may be seen as an unreliable friend, but it is not clear why blame would be an appropriate reaction to my forgetting and for my omission. My view is that here blame is not appropriate.

In order to suggest that dispositional beliefs play an important role in our actions Peels presents the following example:

(...) Imagine that I am teaching a class on evolutionary theory and, in the course of my lecture, tell the students that Darwin's *On the Origin of Species* was published in 1859. This is something I tell them because I believe it, because I take it to be something that the students ought to know, because I believe that there are students in the room, because I believe that I can transfer knowledge to my students by telling them something, and so forth. But (...) I need not consciously consider all these reasons in order for it to be true that my telling the students that Darwin's *On the Origin of Species* was published in 1859 is based on those reasons.¹⁹⁴

Peels claims that those dispositional beliefs play an important role in his reasons for informing the students regarding the year in which *On the Origin of Species* was published. And the same, according to Peels, applies to cases in which our actions are blameworthy or praiseworthy. However, here my response is that although it is true that some dispositional beliefs can explain one's action, contrary to what Zimmerman claims, it is not clear that they make the agent blameworthy or praiseworthy for performing it. For instance, the reason one may lie is because the person they lie to will suffer if they discover the truth. One may believe this dispositionally or occurrently (this can change at different times). If the belief is always only dispositional I do not reflect on it and I do not assess whether this reason makes lying wrong in that case. But when the belief is occurrent I can decide whether to reflect on it and ask myself whether lying is wrong. If I conclude that lying would be wrong, then I can be appropriately blamed if I go ahead and lie. But if I never consider whether preventing some suffering is a good reason to lie, and my

belief is only dispositional, it seems it is not appropriate to blame myself for lying or to be blamed by others. Of course other reactive attitudes might be appropriate, such as frustration, pity or disappointment.

Dispositional beliefs can explain my action but I cannot be aware of them motivating my action, since I cannot include them in my weighing of reasons. Suppose I take an implicit-association test and I find out that I have some harmful implicit bias. Suppose I now consider what to do with this discovery. After some reflection, I then decide to take a training course aiming at undermining my harmful implicit bias. Suppose that taking the training course is the right thing to do. The implicit bias I have is a dispositional belief. It is true that it contributes to explaining my action, that is, it is true that I take the course because I have that dispositional belief. But the reason why I decide to take the training course is not simply because I have an implicit bias, but because I know I have an implicit bias. And this occurrent belief is what I consider when I reflect on whether to attend the course or not. And if I decided not to attend the course although I think it is the right thing to do, I would be blameworthy for not attending the course. But in order to know that attending the course is the right thing to do to get rid of my harmful dispositional belief, it is not sufficient that I have that dispositional belief. I also need to hold the occurrent belief that I have that harmful implicit bias. This suggests that occurrent beliefs are the kind of beliefs relevant to culpability and that dispositional beliefs contribute to explaining an action without being relevant to one's responsibility.

In conclusion, I think Peels fails to successfully criticise Zimmerman's view that only occurrent beliefs are required for culpability and that, consequently, the

Akrasia Thesis is false. First, because Peels does not take into account cases where one is blameless for failing to activate their dispositional beliefs. Secondly, because although it is clear that dispositional beliefs play a role in explaining our actions, Peels fails to show that they play a role in explaining *our culpability* for those actions.

2.0

In this section I will examine some considerations made by George Sher which might be taken as causing problems for the Akrasia Thesis. In his book *Who Knew? Responsibility Without Awareness*, Sher points out that there are some cases of forgetfulness, poor judgement and lack of moral insight which are not cases of akratic action (*e.g.* usually we don't forget to keep a promise because we have performed an earlier action against our better judgement which has caused our forgetting) but assessing them as blameless would require dramatically revising our intuitive judgements of responsibility. Thus, Sher discusses whether there are reasons to hold a view that would entail the non-culpability of such cases. He cannot find reasons to hold such a view. According to the Akrasia Thesis, cases not involving akratic action are blameless. Therefore, Sher's consideration could undermine the Akrasia Thesis.

Sher begins by noting that there are some contemporary philosophers (such as

Zimmerman) who acknowledge that dispositional beliefs explain actions, but since these beliefs are not present to consciousness, these philosophers believe that they are not relevant beliefs to answering the *practical* question of whether we are responsible for our actions.¹⁹⁵ For these philosophers, Sher claims, the concept of responsibility has its origin in the mental states of the agent when she is deciding what she should do (which is also the time when she decides whether to act according to or against her better judgement). The conditions of applications of responsibility as a practical concept may reflect this origin. If responsibility were indeed a practical concept, then there might be strong support for the claim that only occurrent beliefs are required for culpability.¹⁹⁶ However, Sher finds it problematic to think of the concept of responsibility as a practical concept.¹⁹⁷ Although it is true that I can only deliberate about my future actions, Sher notes that the concept of responsibility is not restricted to the first person and is not only oriented to the future. Blame and praise for an act are appropriate only when the act is completed, in the past. Also, blame and praise are more often directed to other people than to ourselves. The concept of responsibility seems then different from the one originating in our deliberations.¹⁹⁸ Moreover, a practical concept of responsibility must be intersubjective and transtemporal, because practical reason requires reciprocity among agents that see themselves as rational among other rational agents. Sher concedes that people cannot see themselves as responsible for some features of

195 Sher (2009), p. 43.

196 *Ibid.*, p. 44.

197 According to Sher, saying that responsibility is a practical concept means that being responsible is not seen as something that describes the agent or explains his action. Being responsible is not seen as a fact about the agent. If responsibility is a practical concept, attributions of responsibility are then about the agent's attempt to decide what are the reasons he has to act.

198 Sher (2009), p. 47.

their acts if they are not (or they do not later become) aware of them. The lack of cognitive connection between the agent's occurrent beliefs and her act makes it impossible for her to regard herself as responsible. However, Sher thinks we do not have reasons to think, as judges, that this limitation on the agent's awareness should be imported into our own different point of view.¹⁹⁹ In other words, Sher claims that if we are aware of something of which the agent is not aware while she deliberates, and because of that she performs a wrong action, it is appropriate to blame the agent, from our perspective.

I think there are problems with Sher's argument. I agree with Sher that blame and praise for an act are appropriate only when the act is completed and that we direct blame and praise to others more often than to ourselves. I also agree that responsibility is not restricted to the first person. However, this does not entail that the mental states of the agent are not relevant to her responsibility. It seems to me that when we judge whether an agent deserves blame for lying, for instance, our verdict does not simply depend on the fact that performing that action was wrong and that we can see that what she did was wrong, but also on whether when she decided to lie she had the belief that her lying would be wrong. However, it does not seem appropriate to blame her if she thought that lying would be the morally right thing to do. After all, we need to be able to distinguish appropriately, in terms of blameworthiness, between those agents who perform wrong actions just because they happen to have false beliefs for which they are not culpable, and those who perform wrong actions knowing that what they are doing is wrong. Therefore, contrary to Sher's claims, the wrongness of the action as recognised by the judge is not the only

factor to consider to determine whether an agent deserves blame. The mental states of the agent (in particular her belief about the moral status of the action) is also essential to draw the line between the bad blameless agent and the bad blameworthy agent.

If the agent who performs an action thinking that it is wrong is mistaken in her belief, but we believe that the action is right, because of this we should not blame her for performing that action. Still, we might want to blame her for acting against her better judgement. Now, as it happens, we may be mistaken too in believing that that action is right. In that case, someone else (someone wiser than us, God, or what have you) may see that the action the agent performed was in fact wrong and will appropriately blame her because of this and also because when the agent performed that action she thought she was performing a wrong action.²⁰⁰ Thus, blame is appropriate only if we combine considerations about the limitation on the agent's awareness with an external point of view that knows about the moral status of the action in that situation (*e.g.* it knows that it is the action that brings about the best consequences). Thus, when we assess whether someone is blameworthy, we need to take into account both her perspective as an agent with limited awareness of the morally relevant features of her action, and whether she actually performed a wrong action (in a non-perspectival sense), being aware that we too could be mistaken about it.

The question of whether an agent is blameworthy is then not only practical in

200

Note that the agent might think that her action is morally impermissible for the wrong reason. I will not explore this issue here, but it seems reasonable to think that an agent may deserve blame even if she does not hold the correct moral theory for why her action is morally wrong. However, an agent could be blameless if the reason why she thinks her action is wrong makes absolutely no sense (*e.g.* if she thought that lying is wrong because Paris is the capital of France), since if that were the case we would probably be reluctant to consider her a moral agent at all.

the sense that it is answered within the agent's deliberation, but also metaphysical, in the sense that an agent is blameworthy when certain conditions are met and these conditions are both about the mental states of the agent and about the moral status of the action. We may never be able to find out whether an agent is blameworthy – either because we are not sure about his mental state at the time of the action or because we are not sure whether the action he has performed is wrong. But suppose it is possible for someone to obtain this knowledge with certainty. That someone could appropriately blame the agent who performed the wrong action believing he was doing wrong.

Sher continues by claiming that in judgements of responsibility the judge should take into account considerations the defendant failed to consider, for instance the fact that the defendant's action was wrong.²⁰¹ This means that the identification of the judge with the defendant is partial. There is then a gap between the first-person perspective of the defendant and the third-person perspective of the judge. Therefore, blaming does not require full identification with the agent, because if it did, we would regard what the agent takes herself to have most reason to do as what she in fact had most reason to do. And so we would not be able to evaluate whether she has attached too little or too much weight to the relevant moral considerations. This means that we would not believe that she has done anything wrong and that she deserves blame.²⁰²

I think Sher's main mistake is the following: his considerations seem to target only views that model culpability entirely on facts accessible from the first-person

201 Note that the defendant can later become aware of the same considerations of which the judge is aware and be the judge of her past self.

202 Sher (2009), p. 63.

perspective. In doing so, Sher overlooks views such as mine which define blameworthiness both in terms of the perspective of the agent and in terms of the action's moral status (in a non-perspectival sense), and still hold that culpability always involves akrasia (the Akrasia Thesis). According to my view, if we do not consider the perspective of the agent, it is likely that we will end up blaming the agent unfairly. As moral agents we all have different moral beliefs which we have acquired throughout our life. I think it is safe to say that our most fundamental moral beliefs are not up to us. We rather acquire them through education. Depending on our environment, we may learn that harming someone is morally acceptable or unacceptable, *e.g.* in order to realise a higher good. Our differences in terms of these fundamental moral beliefs are not dependent on our decisions. What is up to us is rather what we choose to do with the beliefs we have. Some people might simply be unlucky with respect to their moral education. Suppose that a person was taught that violence is most of the time a legitimate means to reach a goal (let us assume that this moral belief is false). It was probably not up to this person whether she formed this belief. However, it might be up to her what to do with that belief. This is because, intuitively, we do not seem to be bound to act according to our beliefs all the time. An agent may see that doing x is wrong but go ahead and perform x regardless, against her better judgement. And it seems appropriate to blame the agent for performing x in this case, if x is indeed wrong in a non-perspectival sense. But if the agent thought that performing x was the right thing to do, it may be inappropriate to blame her for performing x , because it may not be up to her that she holds the belief that performing x is what she should do. And if it is not up to her that she holds

the false belief that she should perform x , then we cannot expect her to have a different belief and blame her simply because, from our perspective, we see that the belief she holds is false.²⁰³

However, this does not mean that we have to drop our third-person perspective and identify completely with the agent, as Sher seems to conceive as the only other alternative. If we blame the agent (or if the agent blames herself) it is because there is also a third-person perspective playing a role in the assessment of the case. There is then an independent factor to consider, a factor external to the agent's mental state. This independent factor is the moral status of the action. If the agent thinks x is wrong as he acts, but x is not in fact wrong, in a non-perspectival sense, it would be unfitting to blame the agent *for performing* x . We might still want to blame her for the fact that she acted against her better judgement that x is wrong, but it would be inappropriate to blame her for the act itself, if this is not wrong. We do not want to blame an agent for doing something morally right or something neutral, that is for doing something that is not impermissible. Of course we might be mistaken about what is wrong and what is right when we evaluate the agent's situation. If so, when I blame for a permissible action which I mistakenly take to be wrong – my expression of blame is inappropriate. In order to blame appropriately we then need to see both through the mental state of the agent and through the moral status of the action. It might even be the case that I am always mistaken about the moral status of the action or about what was the mental state of the agent at the time of the action. My blaming will then always be inappropriate. Sometimes we might be

203

My underlying assumption here is that it is inappropriate to blame someone for acting according to their better judgement.

right in judging the moral status of the action and we might see something morally relevant that the agent cannot see. And that makes us more morally insightful, but it is unfair (if not presumptuous) to use this superiority (we think we have) to conclude that it is appropriate to blame the person who is not aware of a certain moral claim we believe to be true. It is not fair to blame that person *just because* she does not hold true moral beliefs and because she acts according to false ones, as Sher seems to suggest. We cannot assume that it is nevertheless her fault that she holds those false moral beliefs. In order for the attribution of blame to be fair, that person must also have chosen to act against her better judgement that it was wrong to perform that action.²⁰⁴

The Akrasia Thesis implies that an agent must be aware of a certain demand to act akratically against it and be therefore an appropriate object of blame. Sher points out that we are constantly influenced by unconscious demands.

An agent responds unthinkingly to a moral demand whenever, as a conscientious person, he automatically does his duty; whenever, as an honest person, he does not even consider lying to extricate himself from a difficult situation; and whenever, as a kindly soul, he unthinkingly steers the conversation away from a hurtful topic.²⁰⁵

Sher wants to suggest that even though an agent does not consciously register the reason-giving aspects of his situation, he can still fulfil the demand. Hence, it could be reasonable for us to retrospectively endorse that demand and to

204

the action is not akratic but it is culpable because of an earlier akratic action.

205

Sher (2009), pp. 67-68.

blame the agent for not fulfilling that demand, even if the agent is unaware of the facts that base that demand.²⁰⁶

I agree with Sher that we are constantly influenced by unconscious demands. However, the claim I dispute is that when we respond unthinkingly to a moral demand and do the right thing automatically we deserve moral praise strictly speaking.²⁰⁷ It seems to me that, from a third-person perspective, if I always do the right thing automatically I may be definitely admired for my natural moral conduct, and I might even represent a moral exemplar. My ability to do the right thing may be appreciated. But this seems different from deserving moral praise as an agent. To deserve praise I must have done something which is the result of a choice over which I have control, and I must have reflected on the moral status of my action. Praise cannot be truly deserved if, for instance, I am not aware that acting otherwise would be wrong (*e.g.* if I tell the truth without knowing that lying is wrong). Hence, in order to be morally praiseworthy for my action I need to contemplate at least two options, conclude which action I should perform, and perform it. Responding automatically to a moral situation by performing the right action only shows that the agent is capable of doing the right thing. It might even suggest that the agent was educated in a morally appropriate way and that she probably would act the same way in similar situations. But it does not show that she deserves praise for that specific action, since the alternative action was never considered. Praise is then deserved when I do the right thing regardless of my possibility to do wrong.²⁰⁸

206

Ibidem.

207

moral responsibility.

See chapter 6 for more discussion on the kind of awareness required for

208

What I mean by the alternative action being considered is *not* that in order to be praiseworthy the agent must have thought, about the wrong alternative, "Or maybe I could do that", but simply that she must have acknowledged its possibility, as something that a bad person

In conclusion, it seems to me that Sher's considerations do not undermine that Akrasia Thesis, since his criticisms only target views that call for a full identification of the judge with the agent. My view combines considerations about the mental states of the agent at the time of the action with considerations about the moral status of the action. As such, my view is not as problematic as the views of responsibility to which Sher objects, and it is compatible with the Akrasia Thesis.

3.0

In this chapter I have undermined some objections to the Akrasia Thesis. After presenting, in 1.0, Zimmerman's view on culpability and ignorance (which implies that culpability always involves akrasia), in 1.1 and 1.2 I have undermined the objections to Zimmerman's view raised by Montmarquet and Peels, respectively. I have shown that the problem with Montmarquet's objection is that he fails to consider the possibility of an agent being open to truth-and value-related considerations and, at the same time, blameless for not asking herself whether what she is doing is wrong. On the other hand, I have argued that the problem with Peels' objection is that he ignores cases where one is blameless for failing to activate one's dispositional beliefs, and I have disputed the claim that dispositional beliefs play a role in explaining culpability for actions. Finally, in 2.0, I have examined Sher's considerations which might seem to cause problems for the Akrasia Thesis. I have suggested that these considerations do not undermine the Akrasia Thesis. The main reason is that Sher's criticisms target views which call for a full identification of the judge with the agent, while there are views could do, for instance.

compatible with the Akrasia Thesis, such as mine, which combine considerations about the agent's mental states with considerations about the moral status of the action.

Chapter 6 – Moral Responsibility and Consciousness

In this final chapter I will investigate what it is that the agent must be aware of in order to be morally responsible and what type of awareness or consciousness²⁰⁹ is required for moral responsibility. Once again, I will focus in particular on moral culpability. If the Akrasia Thesis is true – as I suggested in the previous chapter – and an akratic act is one the agent chooses to perform despite her better judgement that she should not perform it, then in order to be culpable for some act clearly the agent must have some awareness of her action.²¹⁰

Michael Zimmerman claims that what the agent must be aware of is the wrongness of the action. More specifically, according to Zimmerman, an agent is directly culpable for an action only if she adverts to the wrongness of the action. I agree with this claim. However, Zimmerman allows possible exceptions to this view. In section 1.0 I will discuss these possible exceptions, namely cases of routine or habitual behaviour. I will examine one case of routine behaviour and one case of habitual behaviour and I will argue that we can accurately assess the responsibility of the agent in both cases, without rejecting Zimmerman's view. Hence, I will conclude that cases of routine and habitual behaviour do not constitute exceptions to Zimmerman's view.

In the following sections I will discuss what type of consciousness is required for moral responsibility. In particular, in 2.0 I will examine how Neil Levy comes to

209

I will use 'awareness' and 'consciousness' interchangeably.

210

I leave aside the empirical question about the frequency with which we act akratically. Zimmerman (1997) and Rosen (2004) suggest that akrasia occurs very rarely, and that we are therefore morally responsible for our actions much less frequently than we might think. I have no issue with conceding that this might be the case.

formulate his Consciousness Thesis according to which in order for an agent to be morally responsible for an action, it is necessary (and sufficient when the other conditions are met) that some facts she takes to be playing a role in explaining why the action is good or bad be *personally available* to her.

In 2.1 I will undermine Levy's claim that personal availability of mental contents is a kind of awareness that is necessary (and sufficient when the other conditions are met) for moral responsibility, suggesting that Levy does not provide good reasons to deny that some occurrent tokening before the action is necessary for responsibility. Finally, in 2.2 I will argue that facts explaining the action's goodness or badness may not be decisive to assess the agent's responsibility. I will instead suggest that, at least as far as blameworthiness is concerned, the kind of mental content that is ultimately decisive to establish whether the agent deserves blame is her belief that the action is wrong, or at least the belief that she should not perform that action. I will then conclude by proposing my own formulation of the Consciousness Thesis according to which in order for an agent to be blameworthy for an action, she must occurrently token – at some point before the action – the belief that the action is wrong, or at least the belief that she should not perform the action.

1.0

In his book *Living with Uncertainty: The Moral Significance of Ignorance*, Michael Zimmerman claims that an agent is *directly* culpable for an action only if the agent believes she is doing something morally wrong in performing that action. In

particular, Zimmerman speaks of this belief in terms of the agent advertent to the wrongness of the act. In order to be culpable the agent needs to contemplate this belief about the moral status of the act. In other words, Zimmerman maintains that the agent is directly culpable only if she holds the belief concerning wrongdoing consciously before her mind. The reason for this is that only if the agent holds the belief concerning wrongdoing before her mind can she act with the intention to heed or not to heed her belief. If the agent does not hold that belief before her mind she cannot deliberately act against it. Thus, Zimmerman seems to maintain that an agent is directly culpable only if her belief concerning wrongdoing plays a role in the reason the agent takes herself to have for acting.²¹¹

I largely agree with this view. However, Zimmerman mentions a possible exception to it. The exception concerns cases of routine or habitual behaviour. In these cases, Zimmerman claims that the agent could be directly culpable for wrongdoing that is deliberate but inadvertent. Zimmerman is thinking of cases where the agent once adverted to the wrongness of her behaviour, but now engages in it routinely, unthinkingly. It might seem that in such cases the intention not to heed the moral reasons against the agent's behaviour persists.²¹² I think that even if it is true that the intention does persist in cases of routine or habitual behaviour, this will not be sufficient to make the behaviour directly culpable. The reason for this will hopefully become clear in what follows, as I am going to argue that these cases do not represent an exception to Zimmerman's view.²¹³

211 Zimmerman (2008), p. 191.

212 *Ibidem*.

213 Zimmerman also claims that an agent can be blameworthy for an action even if this action is not in fact wrong. I have discussed in previous chapters why I reject this claim.

In his book Zimmerman does not examine any example of routine or habitual behaviour. It is not clear whether Zimmerman intends a distinction between ‘routine’ and ‘habitual’. I will consider them as distinct terms. First, I will turn to consider a case of routine behaviour in order to show that it does not constitute an exception to the view that for an agent to be directly culpable for an action, she must advert to the wrongness of her act. Then I will do the same with a case of habitual action.

I take routine behaviour to be a type of behaviour performed according to some regular procedure. Here is an example of routine behaviour:

Suppose I am unemployed. I have been applying for jobs for a long time with no luck and I am running out of savings. I finally receive a call for a job as a sales representative. I have never worked in sales. To my surprise I get the job and I start a period of training. During this period I am told that in order to close a deal I must follow a procedure which includes telling costumers a number of lies: that I can get them a discount on the product, that we are running out of stock so they should act now, that the competition is rubbish, that in the last year I have massively increased my customer satisfaction rating, and so on. As I am exposed to this training, I come to the conclusion that telling costumers lies would be wrong for me to do. However, since I am broke and I need this job, I succumb to my self-interest and I agree to follow this procedure. When I start working I am very aware that every time I tell the costumer a lie I am doing something wrong, to the point that I get very nervous and I feel guilty at the end of each interaction with a costumer. However, after a few months, mine becomes a routine behaviour. I simply follow the procedure and I do not think about the moral status of my behaviour. I stop feeling guilty and my nervousness fades away. This allows me to carry out my work efficiently.

Suppose that the behaviour described in the example – my telling lies to customers – is in fact, in a non-perspectival sense, morally wrong. Zimmerman would probably say about this case that I might be directly culpable for telling lies to the customers even when I do it as a routine action, that is, even when I do not advert to the wrongness of that action. However, to me this seems mistaken. The reason why I think this case does not represent an exception to Zimmerman's view is that it can be explained by applying the distinction between direct and indirect responsibility. If this is a case of culpability, it is a case of indirect culpability. Zimmerman claims that in order to be indirectly culpable for an action y , an agent must be directly culpable for x , where y is a consequence of x . Moreover, he claims that, as a second condition of culpability, there must be a cognitive connection between x and y , that is, the agent must foresee that performing x will (or might) bring about a wrong action y . As far as I'm concerned, I do not see why my example could not be accurately assessed according to these conditions. I will now explain how I think the example can be correctly assessed. In order to do that I will need to distinguish three versions of the case, for it is still – for my purposes – under-described.

In the first version, when I advert to the wrongness of telling lies to customers, during my training, I foresee that continuing to work for that company will turn my wrong behaviour into a *routine* wrong behaviour. I predict that I might end up lying to customers without even thinking of whether what I am doing is wrong. Being aware of this risk, I decide to go ahead and continue working for this company. This makes me culpable – though indirectly – for lying to customers *routinely*. My lying routinely is a consequence of my decision to continue working

for the company and these actions are cognitively connected by the fact that I know that I might end up telling lies to customers *routinely*, and that this will prevent me from reflecting on the moral status of my action.

Now consider a second version of the case. Suppose that during my training I do not contemplate – at any time – the idea that continuing to work for that company might put me in a position where I start telling lies to customers routinely, without asking myself whether what I am doing is wrong. It seems to me that in this version of the case I am not culpable – either directly or indirectly – for routinely telling lies. The reason for this is that, first, I am not adverting to the wrongness of that routine behaviour at the time of the action, and, secondly, I have never thought that I will end up performing such routine wrong actions.

Before stating my conclusion, I would like to consider a third version of the case. During my training I realise that this job may lead people to lie to customers routinely without considering whether it is wrong to do so. However, I also believe that this will not happen to me. Actually I believe that I would be able to stop lying to customers at any point just because it is wrong, and to go back to behaving (morally) well. Even in such a case of self-overestimation it seems to me that I would be neither directly nor indirectly culpable for routinely telling lies to customers. After all, at the time when I adverted to the wrongness of the behaviour, I did not believe *I* would end up performing those wrong actions routinely. In conclusion, I disagree that cases of routine actions may represent an exception to Zimmerman's view of culpability, since I have shown that either they might fall under cases of indirect culpability or they might be cases of blameless behaviour.

In addition to routine actions, Zimmerman also indicates habitual actions as a

possible exception to his view that in order to be directly culpable for an action, the agent must advert to its wrongness. Whereas routine actions are performed following some regular procedure, I take habitual behaviour to be behaviour that is typical – without following a particular procedure – and that is more difficult to stop than routine behaviour.²¹⁴ I will now turn to consider a case of habitual behaviour in order to show that, as well as routine actions, habitual actions should not be seen as a possible exception to Zimmerman's view. The example I am going to assess is the following:

Suppose I move into a new house. I rent the property together with three house-mates who are strangers to me. My bedroom is on the ground floor. The first day I wake up in the new house I realise that my room is stuffy, so I decide to open the window. I then take a shower, have breakfast and get ready to go to work. My house-mates are already at work. I am about to leave when I realise that, although my bedroom window has been open for a while, the room is still stuffy. Leaving the window open for several hours while I am at work would ensure that I find a well-ventilated room when I come back from work. However, leaving the window open when nobody is home would expose the house to the risk of burglary. I could easily carry all my valuables with me to work, but I realise that leaving the window open would put my house-mates' valuables at risk too, since my bedroom door cannot be locked. Regardless, I decide to leave the window open and go to work, without caring about my house-mates' belongings. I do this every day until leaving my bedroom window open becomes a habit. Once this has become a piece of habitual behaviour, I cease to think about whether leaving my window open all day puts my house-mates' valuables at risk. I stop thinking about whether that action is wrong.

214 I set aside any more precise characterisation of the difference between 'routine' and 'habitual', hoping that the difference will intuitively emerge after having considered examples of both kinds of behaviour.

Now suppose that my action is in fact wrong, in a non-perspectival sense, and that I should not leave my bedroom window open all day because it puts my house-mates' belongings at risk of being stolen. Again, I assume Zimmerman would claim that this case could represent an exception to his view of direct culpability, the view that in order for an agent to be directly culpable for an action she must advert to the wrongness of the action. However, I do not see why this would be considered an exception, since I think this case may be correctly assessed – like the routine action case – either as a case of indirect culpability or as a case of non-culpability. When I decide to leave the window open thinking that it is wrong to do so, I am directly culpable for that action. However, when I start doing so from habit, without thinking about the moral status of the action, there are, again, at least three possible scenarios to consider.

The first scenario is one in which at some point before I start leaving the window open habitually I come to realise that I will (or might) develop the habit of performing that wrong action. In this scenario, since I have foreseen that I might start leaving the window open habitually (without thinking about the action's wrongness), and since my continuing to leave the window open will consequently make my behaviour become habitual, I am indirectly culpable for the resulting habitual wrong actions.

The second scenario is one in which I never consider that my wrong action of leaving the window open could become habitual and that I might stop thinking about whether it is wrong to perform it. In this alternative scenario I would be blameless for habitually leaving the window of my bedroom open all day, since I have never adverted to the possibility of that wrong action becoming habitual, and to the possibility that I will stop

thinking about its moral status, as something that will lead me to continue performing that wrong action.

As in the case of routine actions, there is at least one more scenario worthy to be considered. This is a scenario in which when I leave the window open the first time, I do realise that this action might become habitual but I believe that that will not happen to me. This is a scenario in which I overestimate my ability to be aware of the moral status of the action at all times. It seems to me that even in this case I would be blameless for performing the wrong action of leaving the window open when it becomes habitual (granted that I am blameless for my overestimation).

Therefore, I disagree with Zimmerman that cases of habitual actions may represent an exception to his view of direct culpability. These actions may simply turn out to be either indirectly culpable or blameless, if we consider the cognitive connection between them and the earlier fully aware actions. In conclusion, I believe that neither routine actions nor habitual actions could represent an exception to Zimmerman's view of direct culpability. My disagreement with Zimmerman concerns whether direct culpability for a particular action requires adverting to the wrongness *of that particular action*. Thus, my view is that **an agent is directly culpable for a particular action only if she adverts to the wrongness of that particular action**. In cases of routine or habitual behaviour, Zimmerman would instead claim that the agent might be directly culpable even though she does not advert to the wrongness of the particular routine or habitual action.

In the next section I will evaluate Neil Levy's view of direct responsibility. Like Zimmerman, Levy thinks that some consciousness concerning an action's moral significance is required for culpability, but not necessarily consciousness of the wrongness

of the action.²¹⁵ Moreover, unlike Zimmerman, Levy does not think that the mental content the agent must be conscious of in order to be morally responsible must be one that the agent contemplates before her mind.

2.0

In his book *Consciousness and Moral Responsibility*, Neil Levy defends the thesis that an agent is morally responsible for an action only if some facts she takes to be playing a role in explaining why the action is good or bad are at least *personally available* to her (*Consciousness Thesis*). This is a necessary condition of moral responsibility. Levy claims that personal availability is sufficient for moral responsibility, assuming that the other conditions of moral responsibility are met. Personal availability does not require occurrent tokening. I disagree with Levy's formulation of the Consciousness Thesis and I am going to argue that some occurrent tokening before the action is required for moral responsibility. Personal availability does not entail occurrent tokening, hence I am going to dispute the claim that personal availability is necessary for moral responsibility (and sufficient when the other conditions are met).

Levy states that, broadly speaking, the consciousness that is pertinent to the debate about moral responsibility is consciousness as "a state with an informational content"²¹⁶ (as distinct from a phenomenal content). In order to be conscious, in this sense, the agent must be aware of this informational content. In general, the consciousness at issue when we talk about moral responsibility is then *awareness*. According to Levy, an agent is aware of something

215 Note that this is not meant to be factive, since the agent's belief about the moral status of the action may be false.

216 Levy (2014), p. 30.

only if the information is both *currently active* and *dispositionally available*. A content is dispositionally available when it is available to report. The same content is also currently active when it plays some role in the agent's cognition, that is when it has an effect on her cognitive systems.²¹⁷

Levy wants to identify the type of conscious content that grounds moral responsibility. Levy claims that it is not plausible that agents are morally responsible for their actions only if they are *occurrently aware* of the significance of those actions or of the reasons for performing them. The reason why this is implausible is that this type of awareness is too demanding – Levy maintains. He presents the following example in order to support his claim:

Suppose that Dr. No intends to kill James Bond using an elaborate machine he has designed. Operating the machinery might require so much concentration that when Dr. No implements his plan he is utterly absorbed in the technical details and has no occurrent awareness of the end toward which he is working.²¹⁸

Levy claims that Dr. No may be responsible for trying to kill James Bond and, if so, he would also be responsible for killing Bond if his plan were successful. Levy concludes that the kind of awareness that grounds moral responsibility need *not* be occurrent tokening at the time of the action. Actually, Levy's claim is even stronger. He believes that moral responsibility does not require occurrent tokening *at any moment*. I will dispute this claim as I think Levy does not provide good reasons to deny that some occurrent tokening is required at some point before the action.

217

Ibid., p. 31.

218

Ibidem.

Levy rightly points out that mere *dispositional awareness* is not sufficient for moral responsibility. For an agent to be dispositionally aware of some content, she must be able to occurrently token that content if she were given the right cues. Dr. No *is* indeed dispositionally aware of his end to kill James Bond.²¹⁹ Levy notes that among all our *dispositionally available* contents, there are some that come to mind spontaneously and others that require a lot of effort to be retrieved. Levy claims that there is a correlation between the degree of accessibility of my mental contents and the degree of responsibility for my failure to utilise them. In order to support this claim he asks the reader to think of dementia patients: “At least in the earlier stages of the disease, the content of a particular memory may still be dispositionally available to them, but it would require more effort on their part or more cues in the environment to retrieve it than is ordinarily the case”.²²⁰ Levy believes that the fact that more effort and cues are needed to recall some informational content is what diminishes the responsibility of the agent for their not recalling that content and then utilising it. Agents such as patients affected by dementia are dispositionally aware of certain mental contents, but they cannot recall and utilise that information and this is why, according to Levy, dispositional awareness is not sufficient for moral responsibility.

For an agent to be responsible for not recalling and utilising some mental content it is then *not* sufficient that the content be merely dispositionally available. The content must also be *online*. This means that it must play some role in guiding the behaviour of the agent.²²¹ This is the case of Dr. No, Levy maintains, since the

219

Ibid., pp. 32-33.

220

Ibid., p. 33.

221

Note that this is not the same as being currently active which means –

goal of Dr. No is to kill James Bond and this guides the series of actions in which Dr. No engages. To further explain the concept of online information – Levy discusses the following example:

Consider Emily, who forgets that today is her wedding anniversary, and therefore makes plans to go out with friends. Nevertheless, the knowledge that it is her anniversary guides some of her behavior: without her realizing the reason for it, she might choose to wear the necklace her husband gave her on their last anniversary, and choose it because it is their anniversary.²²²

Levy claims that, in the case of Emily, even though it is true that her knowing that it is her anniversary guides her behaviour when she chooses what necklace to wear, this, by itself, does not show that she is responsible for forgetting her anniversary and making other plans. If we imagine an alternative version of the case in which Emily also suffers from dementia, although the content that it is her anniversary is online, she might still be *not* responsible for failing to utilise that content in her planning, because it is not available *enough* to her.²²³

Levy finally claims that the kind of contents that ground moral responsibility are those contents that are *personally available* to the agent. This means personally available for report and for the driving of further behaviour. Note that in order to ground responsibility these contents need not to be personally available for report under any condition though, but only under report-conducive conditions, according

according to Levy – having some effect on the agent's cognitive system.

222 Levy (2014), pp. 33-34.

223 *Ibid.*, p. 34.

to Levy.²²⁴ Information is personally available to the agent when it is online *and* when it can be easily and effortlessly retrieved by the agent so that she can utilise it in reasoning.²²⁵ Thus, all personally available mental contents are also online contents, but some online contents may be personally *unavailable* to the agent. Levy explains that the information must be online in order to be personally available because we sometimes become aware of some content by retrieving it. If the information is not already online before its retrieval, then it will not count as personally available until it is actually retrieved. An agent can retrieve a content easily and effortlessly when she would need only ordinary cues and no special prompting to retrieve that content. An example of special prompting is asking leading questions. Levy points out that usually when an agent is able to easily and effortlessly retrieve a content it is because that content is *intermittently occurrently tokened*. This means that the agent often entertains that content.²²⁶ I believe this claim makes Levy's position weaker, as I will argue in the next section.

In conclusion, Levy's position is that "agents need to be conscious of certain of their mental states in order to be (directly) morally responsible for their actions".²²⁷ Some type of consciousness is then necessary. And the type of consciousness which Levy believes to be sufficient for moral responsibility, assuming that the other conditions are met, is personal availability.

Finally, Levy moves on to discuss what it is that the agent must be conscious of

224
225
availability.
226
227

Ibid., p. 32.
I take this to be a necessary and sufficient condition of personal
Ibid., pp. 34-35.
Ibid., p. 35.

to be directly morally responsible for her action. He claims that there are certain facts that play a role in explaining why responsibility is valenced positively or negatively, that is why an agent is praiseworthy or blameworthy. These facts are, for example, what makes an action good or bad (*e.g.* the fact that it benefits or harms someone). But the facts that the valence of direct moral responsibility really tracks – according to Levy – are not these actual facts, but those facts the agent *takes to be* relevant. I am generally sympathetic with this position, though I will later dispute its details. In order to make his point clearer, Levy provides the two following examples:

An agent may be morally blameworthy for what he takes to be a theft from an impoverished widow, even if it turns out that the woman is fabulously wealthy and the bag he sneaks off with has in fact been discarded by her. Similarly, agents are not morally praiseworthy for donations to good causes if they mistook the tin into which they dropped the cash for a collection box for ‘People for cannibalism’.²²⁸

Levy reformulates the consciousness thesis by adding that in order to be morally responsible for an action the agent must be conscious of the facts she takes to be playing a role in explaining the moral valence of the action.²²⁹ And it is sufficient that the facts the agent takes to be playing a role in explaining why the action is good or bad be personally available to her. In order to be praiseworthy/blameworthy for an action the agent must be conscious of some of the facts (under some appropriate description) which she takes to be those which make

228

Ibid., pp. 37-38.

229

'Being conscious of the facts' is not factive here, since what Levy is describing is a perspectival phenomenon.

the action good/bad. In 2.2 I will dispute the claim that the facts that make the action good/bad are ultimately relevant to assess the agent's responsibility. But before that, in 2.1, I will argue that personal availability of mental contents is not the kind of awareness necessary for moral responsibility (and sufficient when the other conditions are met).

2.1

In order to evaluate Levy's view, I will go back to the example of Dr. No and assess the case in light of Levy's formulation of the Consciousness Thesis. According to Levy, in order for Dr. No to be blameworthy for planning to kill James Bond, it is sufficient that some of the facts (under some appropriate description) which Dr. No takes to be those which make killing James Bond a bad action be personally available to Dr. No. I believe that some occurrent tokening before the action is required for responsibility in order for Dr. No to be culpable for planning to kill Bond. Now, if personal availability of a mental content entailed occurrent tokening, when the facts I take to be those which make my action bad are personally available to me, they are at the same time also occurrently tokened. And if this were the case I would not disagree with Levy's condition of moral responsibility. However, it seems to me that there is no such a relation between occurrent tokening and personal availability.

Suppose that Dr. No has personally available to him the belief that Bond's family will be devastated if he kills Bond. Moreover, suppose that the belief that this

is what makes planning to kill Bond bad is also personally available to Dr. No. From this it does not follow that this belief is also occurrently tokened. A mental content is personally available if it is online – that is, it plays a role in guiding the agent's behaviour – and if it can be easily and effortlessly retrieved by the agent and utilised in the agent's reasoning. But a mental content can have these features and still not be occurrently tokened, because it might never be retrieved. The content might play a role in guiding Dr. No's behaviour, for example in the case in which Dr. No planned to kill Bond exactly because he wants to hurt Bond's family. But if Dr. No's mental content (his belief that killing Bond will hurt Bond's family) is merely personally available, although it is true that it can be effortlessly retrieved, it has not been retrieved yet, that is, it is not occurrently tokened. And once it becomes occurrently tokened, it is no longer personally available. Therefore, it seems that personal availability does not entail occurrent tokening.

Let me now say something about why I think that personal availability is not the type of awareness necessary for moral responsibility (and sufficient when the other conditions are met). The main reason has to do with the counterfactual feature of Levy's notion of personal availability. According to Levy, if Dr. No has never been occurrently aware of what he takes to be the reason why his killing Bond is a bad action, he may still be blameworthy for planning to kill Bond, because he would retrieve that mental content if some ordinary cue made him become occurrently aware of it. But this condition seems unfair. The reason is that it seems unfair to hold Dr. No responsible for an action guided by some mental content on the grounds that his counterpart would retrieve it when exposed to some ordinary cue. Similarly, I

think that Levy's claim that there is a correlation between the degree of accessibility of the belief and the degree of responsibility for failing to utilise it is also mistaken. For it does not matter how difficult or easy it is for Dr. No to become occurrently aware of the fact he takes to make his planning to kill Bond a bad action. Dr. No does not have control over whether to retrieve that mental content (he does not know what to retrieve). He does not know that it would be effortless for him to retrieve the belief that Bond's family will be devastated by the loss of Bond. So it would be unfair to blame him when he has no control over whether he is able to retrieve effortlessly that mental content or not. It is only when he becomes occurrently aware of that mental content that he will be able to decide what to do with it: to not hurt Bond's family, because it is bad, or to go ahead with his plan to kill Bond, despite that belief (or because of it). And this is what is decisive to establish whether he will be blameworthy for his choice to continue with his plan.

Levy claims that easy and effortless retrieval of a content means that the agent can retrieve it as a response to ordinary cues, that is when no special prompting is required to retrieve that content. Asking a leading question is considered an example of special prompting. If you notice I am forgetting my friend's birthday, a leading question you might ask me is: 'Aren't you forgetting something today? You might want to call someone today'. According to Levy, a response to an ordinary cue would instead be, for example, that when I see the phone I remember that I should call my friend to wish her happy birthday *because* it is her birthday.²³⁰ However, when we switch to beliefs about the moral status of actions, it is less clear what an ordinary cue would look like. Suppose we ask Dr. No: 'Don't you see that by using this machine on Bond you will hurt other

people too?’ This might be considered a leading question to retrieve Dr. No’s belief that Bond’s family will be devastated by Bond’s death. However, it is much less clear what would be, in this case, an ordinary cue, as it seems implausible that, for example, by looking at his own machine Dr. No would become occurrently aware that some other people will suffer as a result of Bond’s murder.

According to Levy the contents that ground responsibility must not only be easy to retrieve but they must also be online. This means that they must play a role in guiding the agent’s behaviour. However, in most cases blameworthy agents are not motivated by the facts they take to be what makes their action bad. They rather perform the action *regardless of* their mental contents about what makes their action wrong, since they often decide, for example, to follow their self-interest instead, and perform the bad action anyway.

Finally, what I think is the most significant weakness that is apparent in Levy’s argument is that, although he denies that occurrent tokening is necessary for moral responsibility, he acknowledges that agents who have a content online and are able to easily and effortlessly retrieve it can do so because that content is intermittently occurrently tokened. This seems to imply that most of the time some occurrent tokening plays an important role in the agent’s situation. Hence, I don’t see why Levy would want to insist that no occurrent tokening at any particular time is required for moral responsibility. Therefore, while I agree with Levy that agents need to be conscious of certain of their mental states in order to be morally responsible for their actions, I disagree on the sense of ‘conscious’ that is fitting here. In particular, it seems to me that some occurrent tokening at some point before the action is required for the agent to be

morally responsible for that action. In the next section I will suggest that the mental content necessary for blameworthiness is a belief that that action is wrong, or at least a belief that I should not perform that action.

2.2

According to Levy's Consciousness Thesis, in order for an agent to be morally responsible for an action, some facts she takes to be playing a role in explaining why the action is good or bad must be *personally available* to her. I agree with Levy that these facts (which might indeed not be facts if the agent is wrong about them) are relevant to moral responsibility, but only in so far as they are connected to a belief about the moral status of the action. In fact I argue that holding the belief about the moral status of the action may ultimately be sufficient.

Levy gives an example in which an agent steals from someone he takes to be an impoverished widow. Levy claims that the agent will still be blameworthy even if the victim turns out to be a wealthy woman. While I agree with this point, Levy might not agree with me about the reason why the claim is true. It seems to me that the reason why the thief will be still blameworthy is that he thinks that an act of theft is wrong regardless of whether it is against an impoverished widow or against a wealthy woman. But the fact that he does not know that his victim is a wealthy woman makes him blameworthy for a slightly different action. It makes him blameworthy only for stealing from a person and not for stealing from a wealthy woman, because he thought he was stealing from an impoverished widow.²³¹

231

There is a sense in which, since the thief thought his victim was an

Now let us turn to the example of the money donation mentioned by Levy. He claims that the agent would not be praiseworthy for the general action ‘donating to good causes’, if they dropped the cash into a tin which they mistook for a collection box for ‘People for cannibalism’. Now suppose I am the agent in question. It seems to me that whether I would be praiseworthy for donating to good causes depends on both my beliefs and the moral status of the action. The case is then under-described. Suppose that I believe I should donate to a good cause. Suppose donating to a good cause is in fact what I should do. I think that ‘People for cannibalism’ is a good cause. Suppose I end up donating to the Red Cross which is, let’s assume, also a good cause. It seems to me that although I am not praiseworthy for donating to the Red Cross, since I thought I was donating to ‘People for cannibalism’, I am praiseworthy for donating to a good cause, because that is what I thought I was doing. If what we are concerned with is praiseworthiness for donating to good causes, it is irrelevant that the cause I ended up donating to was different from the one *I believed* I was donating to.

Now suppose instead that I believe that I should *not* donate to ‘People for cannibalism’, but I have a morbid desire to see more people being eaten, so I go ahead and drop the cash into the tin, believing I am donating to ‘People for cannibalism’, when I am in fact donating to the Red Cross. In this version of the case I am not praiseworthy for donating to a good cause, because I believed I was not donating to a good cause when I dropped the cash into the tin. I did not know that I

impoverished widow, he would be *more blameworthy* if he was aware that it is morally worse to steal from an impoverished widow rather than from a wealthy woman (assuming that this is true). I will not discuss degrees of responsibility here. My focus is rather on the distinction between blameworthy and blameless cases and on what it is that the agent is responsible for.

was in fact donating to a good cause. But there is a lack of cognitive connection – in Zimmerman’s words – between my decision to donate money to a cause I do not consider good and my action of donating money to a good cause. And this makes me not praiseworthy for donating money to a good cause.

Levy’s formulation of the consciousness thesis states that in order for an agent to be morally responsible for an action, some facts (under some appropriate description) she takes to be playing a role in explaining why the action is good or bad must be personally available to her. I disagree with this formulation of the thesis. In 2.1 I have discussed why I reject personal availability of a content as the kind of awareness that is necessary for moral responsibility (and sufficient when the other conditions are met). However, I also dispute the part of the formulation stating that the agent must take these facts to play a role in explaining why the action is good or bad, as it seems to me that this condition is not what ultimately grounds moral responsibility. The reason for this is that there will be cases in which the agent must decide among a number of actions that seem to him all good or all bad (as in cases of moral dilemmas), but perhaps not all equally so. This is glaring especially when blameworthiness is in question. In cases of moral dilemma it seems reasonable to believe that it is not the case that the agent will be blameworthy regardless of which action he decides to perform. When all the alternatives look bad to the agent, she may be conscious of some facts that make the actions bad but she may still be unaware of the facts that make a certain option worse. This will happen if she is not conscious of the relative weight of the facts that make the actions bad. Therefore, it seems to me that the consciousness thesis should not be simply about the facts that make an action good or bad. It should be rather about some facts that make the

action required or wrong, that is some facts that make the agent conclude that she should or should not perform that action. In some circumstances, though, the agent's situation will not allow for consideration and weighing of different facts (*e.g.* because of the agent's cognitive ability or lack of time to reflect on what decision to make). Still, if the agent takes a certain action to be wrong, if she has an occurrent belief (not simply personally available as Levy claims) that she should not perform that action at some point before its performance, then she will be blameworthy for performing that action, if the action is also wrong (in a non-perspectival sense). Therefore, focussing in particular on blameworthiness, I believe that the consciousness thesis should be revised as follows: in order for an agent to be blameworthy for an action she must occurrently token – at some point before the action – the belief that the action is wrong, or at least the belief that she should not perform the action.

3.0

In conclusion, in this final chapter I have investigated what it is that the agent must be aware of in order to be morally responsible and what type of consciousness is required for moral responsibility, and in particular for moral culpability. In 1.0 I have discussed possible exceptions to Michael Zimmerman's thesis that an agent is directly culpable for an action only if she adverts to the wrongness of the action. These exceptions, suggested by Zimmerman himself, are cases of routine or habitual behaviour. I have examined one case of routine behaviour and one case of habitual behaviour. I have shown that we can accurately assess the agent's culpability in both

cases without rejecting Zimmerman's thesis, since the cases can be seen as either cases of indirect culpability or as cases in which the agent is blameless. I have therefore argued that cases of routine and habitual behaviour do not constitute exceptions to the view that an agent is directly culpable for a particular action only if she adverts to the wrongness of *that particular action*.

In 2.0 I have considered Neil Levy's formulation of his consciousness thesis according to which in order for an agent to be morally responsible for an action, it is necessary (and sufficient when the other conditions are met) that some facts she takes to be playing a role in explaining why the action is good or bad be *personally available* to her. In 2.1 I have undermined the claim that personal availability of mental contents is the kind of awareness that is necessary for moral responsibility (and sufficient when the other conditions are met). I have shown that Levy does not provide good reasons to deny that some occurrent tokening before the action is necessary for moral responsibility. Finally, in 2.2 I have argued that facts explaining the action's goodness or badness may not be decisive to assess the agent's responsibility. I have instead suggested that the kind of mental content that is ultimately decisive to assess an agent's blameworthiness is her belief that the action is wrong, or at least the belief that she should not perform the action. I have therefore concluded by advancing an amended version of the Consciousness Thesis (focused on moral culpability). This claims that in order for an agent to be blameworthy for an action she must occurrently token – at some point before the action – the belief that the action is wrong or, more broadly, the belief that she should not perform the action.

Conclusion

The main contribution of this thesis has been to provide more reasons to undermine objections to volitionism. Most of the objections I have considered come from attributionists and concern the epistemic condition of moral responsibility.

First, I have explained some differences between volitionism and attributionism as the main kinds of account of moral responsibility. I have given sufficient reason in favour of volitionism and against attributionism to motivate further articulation of volitionism (chapter 1). Then I have explored the question of whether and when it is appropriate to blame someone for holding a belief (chapter 2). I have advanced the principle that it is appropriate to blame an agent for a false belief only if she chose not to inform herself and she knew she should, in order to avoid forming a false belief. Next, I have discussed the relation between blameworthiness of the agent and moral status of the action for which she is blameworthy (chapter 3). I have defended from objections the Objective View, which states that an agent is blameworthy for performing an action only if the action is objectively wrong. In addition, I have proposed a view that combines the wrongness of the action with *certain* mental states of the agent. This view states that an agent is blameworthy for performing an action only if (a) she believes that the action is wrong *and* (b) the action is wrong in a non-perspectival sense. In chapter 4 I have defended the Ignorance Thesis from some objections. This is the thesis that whenever an agent acts from ignorance, she is culpable for the act only if she is culpable for the

ignorance from which she acts. I have shown that the Ignorance Thesis applies both to cases of factual ignorance and to cases of moral ignorance. In chapter 5 I have undermined some objections to the Akrasia Thesis which states that moral culpability always involves akrasia. I have suggested that volitionists are right in claiming that an agent is culpable for an action only if that action is either akratic or culpably ignorant. Finally, I have discussed what it is that the agent must be aware of in order to be morally responsible and what type of consciousness is required for moral responsibility (chapter 6). I have undermined the thesis that, in order for an agent to be morally responsible for an action it is necessary (and sufficient when the other conditions are met) that some facts she takes to be playing a role in explaining why the action is good or bad be personally available to her (Consciousness Thesis). I have then advanced an alternative formulation of the Consciousness Thesis which focuses on moral culpability. The thesis states that in order for an agent to be blameworthy for an action she must occurrently token – at some point before the action – the belief that the action is wrong, or at least the belief that she should not perform the action.

Future research on the epistemic condition of moral responsibility could include the following: a positive argument for volitionism; the relation between degrees of belief and degrees of culpability; the risk of infinite regress in tracing; the question about the frequency of akratic acts; and the question of when exactly the agent must advert to the wrongness of her action in order to be blameworthy for it.

Bibliography

- Brandt, Richard. 1958. "Blameworthiness and Obligation", in A.I. Melden (ed.), *Essays in Moral Philosophy*, Seattle: University of Washington Press.
- Dewey, John. 1891. *Outlines of a Critical Theory of Ethics*. Reprint, New York: Hillary House, 1957.
- Feldman, Fred. 1986. *Doing the Best We Can*. Dordrecht, Holland: D. Reidel Publishing Company.
- Fischer, J.M. & Tognazzini, N. 2009. "The Truth about Tracing". *Nous* 43: 531-56.
- FitzPatrick, W.J. 2008. "Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge". *Ethics* 118: 589-613.
- Guerrero, Alexander A. 2007. "Don't Know, Don't Kill: Moral Ignorance, Culpability, and Caution". *Philosophical Studies* 136: 59-97.
- Haji, Ishtiyaque. 1997. "An Epistemic Dimension of Blameworthiness". *Philosophy and Phenomenological Research* 57: 523-544.
- Harman, Elizabeth. 2011. "Does Moral Ignorance Exculpate?" *Ratio* 24: 443-468.
- Jackson, Frank. 1991. "Decision-theoretic Consequentialism and the Nearest and Dearest Objection". *Ethics* 101: 461-482.
- Khoury, Andrew C. 2011. "Blameworthiness and Wrongness". *Journal of Value Inquiry* 45: 135-146.

- King, Matt. 2015. "Epistemic Tracing". *Responsibility – The Epistemic Dimension* Workshop. VU University Amsterdam.
- Levy, Neil. 2005. "The Good, the Bad, and the Blameworthy". *Journal of Ethics & Social Philosophy* 1: 2-16.
- Levy, Neil. 2007. "Doxastic Responsibility". *Synthese* 11: 127-155.
- Levy, Neil. 2009. "Culpable Ignorance and Moral Responsibility: A Reply to FitzPatrick". *Ethics* 119: 729-741.
- Levy, Neil. 2014. *Consciousness and Moral Responsibility*. Oxford: Oxford University Press.
- McCormick, Miriam. 2011. "Taking Control of Belief". *Philosophical Explorations* 14: 169-183.
- Mele, Alfred. 1987. *Irrationality*. New York: Oxford University Press.
- Montmarquet, James. 1999. "Zimmerman on Culpable Ignorance". *Ethics* 109: 842-845.
- Nagel, Thomas. 1979. *Mortal Questions*. Cambridge: Cambridge University Press.
- Owens, D. 2000. *Reason without Freedom: The Problem of Epistemic Normativity*. London: Routledge.
- Peels, Rik. 2011. "Tracing Culpable Ignorance". *Logos and Episteme* 2: 575-82.
- Rosen, Gideon 2003. "Culpability and Ignorance". *Proceedings of the Aristotelian Society* 103: 61-84.

- Rosen, Gideon. 2004. "Skepticism about Moral Responsibility". *Philosophical Perspectives* 18: 295-313.
- Sartorio, Carolina. 2007. "Causation and Responsibility". *Philosophy Compass* 2/5: 749-765.
- Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Sher, George. 2009. *Who Knew? Responsibility without Awareness*. New York: Oxford University Press.
- Sinnott-Armstrong, Walter. 1984. "'Ought' Conversationally Implies 'Can'". *Philosophical Review* 93: 249-261.
- Smith, A.M. 2005. "Responsibility for Attitudes: Activity and Passivity in Mental Life". *Ethics* 115: 236-271.
- Smith, Holly. 1983. "Culpable Ignorance". *The Philosophical Review* 92: 543-571.
- Smith, Holly. 1991. "Varieties of Moral Worth and Moral Credit". *Ethics* 101: 279-303.
- Streumer, Bart. 2003. "Does 'Ought' Conversationally Implicate 'Can'". *European Journal of Philosophy* 11: 219-228.
- Van Woudenberg, R. 2009. "Excusing Conditions for False Beliefs: Ignorance and Force". *American Philosophical Quarterly* 46: 373-386.
- Vargas, M. 2005. "The Trouble with Tracing". *Midwest Studies in Philosophy*, 29: 269-91.

- Watson, Gary. 2004b. *Agency and Answerability: Selected Essays*. Oxford: Clarendon Press.
- Williams, Bernard. 1970. "Deciding to Believe". In Bernard Williams, *Problems of the Self*. Cambridge: Cambridge University Press, pp. 136-151.
- Zimmerman, Michael J. 1997. "Moral Responsibility and Ignorance". *Ethics* 107: 410-426.
- Zimmerman, M. J. 2008. *Living with Uncertainty: The Moral Significance of Ignorance*. Cambridge: Cambridge University Press.