Half-Souled Bodies: Punishing Civil Servants for Past State Atrocities

Ka Long Chan
St Antony's College
M.Phil. in Politics (Political Theory)

Thesis submitted in partial fulfilment of the requirements for the degree of M.Phil. in Politics (Political Theory) in the Department of Politics and International Relations at the University of Oxford.

Trinity term, 2022
Word count: 27025

Half-Souled Bodies: Punishing Civil Servants for Past State Atrocities

Ka Long Chan
St Antony's College
M.Phil. in Politics (Political Theory)

Thesis submitted in partial fulfilment of the requirements for the degree of M.Phil. in Politics (Political Theory) in the Department of Politics and International Relations at the University of Oxford.

Trinity term, 2022
Word count: 27025

Abstract: This thesis investigates the principal question of whether civil servants should be punished for past state atrocities under transitional circumstances. It looks into three sub-questions to answer this principal question. First, it inquires into whether retributivism provides an application justification of punishment in transitional circumstances. By addressing challenges about the effectiveness of punishment, the standing of the punisher, and the retroactivity of punishment, it argues that expressivist retributivism offers a defeasible justification for punishing civil servants in transitional circumstances. Second, it examines how civil servants can be held complicitly responsible for state atrocities. It proposes that an individual is complicitly responsible for a collective wrongdoing if she has taken an action that carries an impermissible risk of causing the collective wrongdoing's success, and this action expresses a culpable choice. It also argues that this principle of complicit responsibility can overcome the problem of overdetermination. Third, it investigates what valid justifications and excuses civil servants can claim to counter blame and punishment. It rejects the cog-theory justification, the superior order justification, and the excuse of absent discriminatory intent. However, it argues that non-culpable ignorance can excuse civil servants' contribution to atrocities, and their ignorance is non-culpable if they have fulfilled their respective level of epistemic obligations. Therefore, this thesis argues that any civil servants who have contributed to state atrocities and cannot appeal to any valid justification or excuse deserve to be punished. If there are no forward-looking considerations that can outweigh retributive justice in the transitional context, these civil servants should be punished.

Chapter 1
Introduction

States sometimes commit atrocities against their people. By atrocities, I mean genocide and crimes against humanity, which refer to a 'widespread or systematic attack against any civilian population' (International Criminal Court, 2011, 3) through murder, extermination, enslavement, deportation, imprisonment, torture, sexual violence, arbitrary persecution, enforced disappearance, apartheid, and other inhumane acts. Atrocities are wrong, so when they have happened, justice demands some responses. When the government that has committed atrocities is still in power, it will likely refuse to admit or address its wrongdoings. However, when the government has undergone a radical change – what I refer to as a transition – due to changes in the regime, leaders, or even boundaries, the transition creates opportunities for recognising and addressing past atrocities. Apart from opportunities, transitions also present unique challenges to addressing past wrongdoings, such as the normalisation of human rights violations, that states normally do not encounter. These opportunities and challenges constitute the transitional circumstances in which just responses occur, and unique moral issues arise.

Due to the scale of state atrocities, state leaders' sheer will and capacity are not sufficient for their successful implementation. Civil servants, therefore, play an indispensable role in enabling and producing state atrocities. The term 'civil servant' acquires different meanings in different contexts. In one sense, we contrast it with leaders who occupy higher ranks in the hierarchical state structure: All state personnel who are below the leadership role are civil servants. In another sense, we contrast 'civil servant' with police and soldiers who perform coercive, 'uncivil' functions. This thesis adopts the former sense. It defines civil servants as state personnel who execute orders from their superiors. Under this definition, police and soldiers who enforce orders from their leaders are also civil servants. Admittedly, civil servants in modern states often give orders, make plans, and exercise discretion – there is seldom a sharp separation between politics and administration as envisioned by Max Weber (2014). However, as we will shortly see, this thesis aims to challenge the predominant focus of international practices on punishing leaders rather than lower-level personnel for past atrocities. By defining civil servants by their distinctively executive nature, I can set up a hard case for examining whether punishment should be extended to civil servants.

The Principal Thesis Question

This principal question of my thesis is this: Under transitional circumstances, should civil servants be punished for past state atrocities? This question has two primary motivations. First, the punishment of individuals is one of the most common responses to past atrocities. The international practice of punishing individuals for atrocities could, at least, be traced to the Nuremberg and Tokyo trials. Both punished individuals for crimes committed during World War II (Isaacs, 2011). The United Nations (UN) includes prosecution initiatives and trials as core components of transitional justice programmes (UN Secretary-General, 2010). It established the International Criminal Tribunal for the Former Yugoslavia (ICTY) and International Criminal Tribunal for Rwanda (ICTR) in 1993 and 1995 respectively (United Nations Security Council, 2021).

In 2002, the International Criminal Court (ICC) began its operation following the Rome Statute. The ICC currently has jurisdiction in 123 member states to prosecute individuals for international crimes (International Criminal Court, 2011; International Criminal Court, 2022). While there are other responses to atrocities, such as reparations and the establishment of truth and reconciliation commissions, the prevalence of punishment as a response calls for critical examination.

Second, while there is an intuitive appeal that civil servants' mass involvement in state atrocities makes them wrongdoers, civil servants have rarely been the targets of punishment. International trials have almost exclusively focused on prosecuting political leaders and high-level cadres (McMahan, 2008; Allais, 2011). For instance, as McCoubrey (2001) notes, the defendants brought to the Nuremberg and Tokyo trials were 'for the most part political leaders of cabinet rank and military officers at very senior levels of command, in most cases at the level of general staff leadership' (289). In contrast, civil servants (including soldiers) often faced no prosecutions or punishment (Malamud-Goti, 1990; McMahan, 2008; Allais, 2011; Drumbl, 2011). Feasibility constraints could partly explain this contrast: National and international courts might have limited resources to bring every civil servant to trial; mass prosecution might instigate conflicts that would reverse the transitional state's precarious peace. Yet, as we will see, the literature has also presented moral arguments that support the conclusion not to punish civil servants. Although the intuition that justice demands the punishment of civil servants may be ill-considered, the apparent conflict still calls for a critical examination of those arguments.

Chapter Sub-Questions

To investigate whether civil servants should be punished for past state atrocities under transitional circumstances, this thesis will tackle three sub-questions, each addressed in a chapter. Chapter 2 will examine this sub-question: Does retributivism justify punishment for past atrocities in transitional circumstances? It will look into Murphy's (2017) argument that the unique challenges of transitional circumstances make retributivism an inapplicable justification of punishment in transitional circumstances. In response, I will argue that retributivism – particularly what I call expressivist retributivism – justifies punishing the perpetrators of past atrocities despite the challenges of transitional circumstances.

Chapter 3 will examine this sub-question: What are the conditions of holding civil servants complicitly responsible for past state atrocities? It will take up the retributivist premise that punishment must be proportional to individual culpability and ask under what conditions can civil servants be held morally responsible for the state's collective wrongdoings. The key challenge addressed by this chapter is the problem of overdetermination, which states that individuals cannot be held responsible for collective wrongdoings when their contribution is causally unnecessary to the collective outcome. I will critically review the conceptions of complicit responsibility proposed by the literature in response to this problem. I will then contend that civil servants are culpable for state atrocities if they have taken actions that carry impermissible risks of causing state atrocities.

Chapter 4 will examine this sub-question: What justifications and excuses can civil servants claim to counter blame and punishment? Accused civil servants often claim

themselves to be ‘half-souled’ – lacking full agency for what they have done: They defend themselves by claiming that they are just cogs in a machine, they are just following orders, they lack discriminatory intent, and they do not know the wrongness of what they have done. Chapter 4 will examine the validity of these defences and, if they are, the conditions of their validity. These conditions will, in turn, guide the determination of whether each civil servant is culpable and deserves punishment.

Methodology and Limitations

The methodology of this thesis will follow what McDermott (2008) describes as the reasoning process of analytical political philosophy: In some particular cases, we are reasonably confident with our (intuitive) judgment about what ought to be done. Then, we use these ‘oughts’ to develop and test normative theories that best capture these ‘oughts’. Next, we use these theories to revise our intuitive judgments and investigate moral questions that we lack a clear answer to. Following this methodology, I will use both real-world cases and thought experiments throughout the thesis to identify intuitive judgments, which will, in turn, confirm or challenge existing theories and ground the development of my theories about punishment in transitional circumstances.

While I am confident that this thesis’s investigation can give insights into the issue of punishing civil servants, there are still three potential limitations. First, like discussions of human rights in the liberal world, I may have adopted intuitions and premises shared only by liberal readers to develop my conclusions. These potential biases may weaken the applicability of my conclusions in transitional societies that may not have a liberal culture. To minimise the effects of these potential biases, I will build my arguments on judgments and principles that I believe are widely shared across the world. Still, I think some biases from my standpoint are inevitable; all I can do is make my assumptions and reasoning process as clear as possible.

Second, the literature on historical injustice and transitional justice sometimes uses the demands of the victims of wrongdoings to support or reject theories about the appropriate responses to wrongdoings. This method either treats the victims as having privileged epistemic access to what justice demands or unconditionally assigns intrinsic value to fulfilling the victims’ demands. I will not use this method in my thesis. It is partly because conducting primary research into atrocity victims’ opinions of civil servants is well beyond the scope of this thesis. It is also because I am not sure about using this method to discover moral truths. As Elster (2004, 97) accounts, ‘vindictive rape’ sometimes occurred in transitional societies as to how the victim population carried out ‘private justice’. For instance, Soviet soldiers, who were the victims of Nazi Germany’s crime of aggression, raped an estimated 2 million German women in the end stages of World War II. I do not think what these soldiers thought was right to do can provide any insights into what justice demands for the people of Nazi Germany.

Third, I understand that punishing civil servants for atrocities is not only a moral but also a legal issue. Legal principles, especially the principles of criminal law, may also answer whether and under what conditions civil servants should be punished in transitional circumstances, and this answer may confirm or contradict the conclusions of this thesis. It is beyond my expertise and the scope of this thesis to investigate the legal dimension of my principal question. However, I argue that the moral reasons for

and against punishing civil servants must be considered in decisions about the punishment of atrocity perpetrators. The decisions to set up courts to put atrocity perpetrators on trial are always political, and I assume that moral reasons are among the many kinds of reasons that political decisions should consider. Moreover, I assume that laws and legal principles should, ultimately, trace moral principles; a legal principle's deviation from the demands of justice constitute a reason against it. Therefore, even though my moral discussions may not offer direct guidance on the practice of punishment after atrocities, they are still crucial to understanding what should be done.

Chapter 2
Retributive Justice in Transitional Circumstances

There are two broad categories of reasons that matter to the appropriateness of punishing atrocity perpetrators in transitional circumstances. The first category includes forward-looking reasons. They concern the benefits and costs that punishment will bring to society. Punishing atrocity perpetrators can bring benefits to society by deterring the future commission of atrocities (Sikkink & Walling, 2007) and supporting the development of the rule of law (European Union, 2015); it can also lead to costs such as preventing the repairment of broken social relations (Tutu, 2009) and instigating violent conflicts (Nino, 1991). I think the decision-makers of transitional societies must consider punishment's effects on people's fundamental interests. For example, if it is almost certain that putting atrocity perpetrators on trial will result in a civil war, decision makers have good reason to grant, at least temporarily, immunity to the perpetrators. However, it is beyond this thesis's normative discussions to examine whether punishing atrocity perpetrators is beneficial or harmful to transitional societies. The effects of punishment depend highly on the transitional context, so there is no general answer to whether the punishment will be supported or rejected by forward-looking reasons.

The second category of reasons that determine punishment's appropriateness is retributivism, a backwards-looking justification of punishment that ignores punishment's consequences. The basic idea of retributivism is that punishment is an intrinsically justified way to address past wrongdoings; retributive justice is achieved through punishment itself. In general, retributivism sets criteria on the proper target, form, and strength of punishment. Retributive justice then serves as a prima facie justification of instances of punishment that meet these requirements. It constitutes a defeasible moral reason for punishment.

State atrocities, including genocide and crimes against humanity, are serious wrongdoings. Based on retributivism, their wrongfulness provides grounds for the justification of punishing atrocity perpetrators in transitional circumstances. This position is supported by Ohlin (2009) and Greenawalt (2013), who identify retributive justice as an indispensable justification of international criminal justice that punishes atrocities.

However, this general position faces serious challenges from Murphy (2017). Although she acknowledges the justificatory force of retributivism in stable democracies, she argues that the challenges of transitional circumstances make retributivism an inapplicable justification of punishment for four reasons: (1) Punishment cannot serve the retributivist goal of vindicating the value of victims; (2) no authority has the standing to punish wrongdoings; (3) punishment for atrocities is ex post facto; and (4) punishment for atrocity perpetrators cannot satisfy the proportionality requirement of retributivism. If these reasons are valid, retributive justice will not constitute any moral reason for punishing atrocity perpetrators in transitional circumstances, and the case for punishing civil servants for past atrocities will be fragile.

Therefore, this chapter aims to investigate this sub-question of the principal thesis question: Does retributivism justify punishment for past atrocities in transitional circumstances?

In the following, I will address Murphy's first three challenges and contend that retributivism is an applicable justification for punishment in transitional circumstances. While the last challenge is also relevant, it will be addressed in the next chapter.

An Expressivist Theory of Retributivism

Before tackling Murphy's challenges, I will first outline the expressivist theory of retributivism. It will provide this thesis's core framework of justification for punishing civil servants for state atrocities. My following outline of expressivist retributivism is owed to Feinberg (1970), Kleinig (1973), and Duff (2001). It is beyond the scope of this thesis to argue for expressivist retributivism's advantages over other retributivist theories. But given the support from the literature on punishment, I suppose that expressivist retributivism cannot be dismissed by just saying 'this is not what retributivism means'.

The first feature of expressivist retributivism is that it justifies punishment through desert claims. Desert claims match certain treatment of a person with the desert bases the person has. Desert bases are usually the person's characteristics or past activities. If the treatment matches the person's desert bases, the person has received what she deserves. Different treatments match with different desert bases. For example, excellency and contribution deserve praise. Here, expressivist retributivism's justification of punishment relies on a specific kind of match – the match between blame and culpability. In other words, punishment is justified by the desert claim that individuals who are culpable for wrongdoings deserve blame. This claim, I believe, is a fundamental moral intuition widely shared across different cultures.

The second feature of expressivist retributivism connects blame with punishment: Punishment is treated as an expression of blame. As Feinberg (1970) defines it, punishment is 'the infliction of hard treatment by an authority on a person for his prior failing in some respect' (95). And in Kleinig's (2011) words, 'It [hard treatment] imposes a substantial cost. The wrongdoer is made to pay in some way' (55). For Feinberg, Kleinig, and Duff, however, the infliction of hard treatment not only imposes high costs on its targets but also expresses blame towards them. Culpable wrongdoers who deserve blame thereby deserve punishment *as* blame.

The third feature of expressivist retributivism is proportionality: Culpable wrongdoers do not just deserve any level of blame; they deserve proportional blame – the severity of blame should match the level of their culpability. The proportionality requirement explains why culpable wrongdoers sometimes deserve hard treatment rather than mere verbal censure. For example, a cheater in a relationship may well deserve condemnation, but a murderer surely deserves much more than that. Because the strength of verbal censure is limited, hard treatment is sometimes needed to express blame severe enough to match with wrongdoers' level of culpability.

The above three features constitute this basic principle of expressivist retributivism: Culpable wrongdoers, and only culpable wrongdoers, deserve proportional punishment as blame. Retributive justice is achieved when instances of this desert claim are satisfied. However, I must note that retributivism only constitutes a defeasible moral reason for punishment. Even

if a person deserves punishment, it is not necessarily true that she *should* be punished – the final 'ought' depends on other moral reasons that may outweigh the desert claims (Feinberg, 1970, 60; Kleinig, 2011, 57). These reasons can be forward-looking considerations based on the effects of punishment; they can also be related to the moral standing of the punisher.

If expressivist retributivism is applicable in transitional circumstances, it will generate a defeasible justification for giving proportional punishment to culpable civil servants as blame. While Murphy does not challenge this conditional statement, she does reject the sufficient condition – she contends that retributivism (not expressivist retributivism but retributivism in general) is not applicable in transitional circumstances. The rest of this chapter addresses three of her reasons for this contention.

Challenge One: Punishment Cannot Achieve Retributive Goals

Murphy's first challenge is that punishing atrocity perpetrators in transitional circumstances cannot achieve retributive goals that can be achieved in stable democracies. While this challenge mainly targets Jean Hampton's theory of vindicative retributivism, I will show that Murphy's arguments fail to challenge the applicability of both expressivist and vindicative retributivism in transitional circumstances. However, I will not argue for expressivist retributivism's advantages over vindicative retributivism within the scope of this thesis.

Vindicative retributivism

I will start by introducing Hampton's (1992) theory of vindicative retributivism. According to this theory, punishing wrongdoers can vindicate the diminished values of the victims, and this vindication is intrinsically good. Hampton defines wrongdoings as actions that constitute diminishment, which are actions that carry the meaning that other people are inferior to the wrongdoer. These actions are typically the treatment of people in ways that violate their entitlements generated by their value; they can also be the propelment of messages that assert the inferiority of others. For Hampton, diminishment is morally problematic not only because its message denies the (Kantian) moral truth that every person is equally valuable but also because it causes moral injuries to its targets: Diminishing actions damage the targets' realisation of value by violating their entitlements and/or the social acknowledgement of their value. In causing moral injuries, wrongdoings not only send a message that denies the truth of moral equality but also shape the world in a way that provides evidence of moral inequality.

According to Hampton, the diminishing effects of wrongdoings generate moral demands for retributive responses that vindicate the diminished value of victims. Because wrongdoings' denial of moral equality is both linguistic and evidential, the retributive responses should communicate a message of moral equality and 'remake the world' (1687) to give evidence to moral equality. Only then can the act of diminishment be annulled, and the diminished value of victims be vindicated. Here, Hampton argues that punishing wrongdoers is 'uniquely suited to do the vindicating, insofar as it created a state of affairs … in which the victim was elevated with respect to the wrongdoer' (1695). By 'defeating' the wrongdoers, punishment communicates a message of equality and provides counterevidence against moral inequality. To conclude, vindicative retributivism's justification of punishment is based on punishment's capacity to achieve the retributive goal of vindicating the victims' diminished value.

Murphy's challenge

Murphy (2017) argues that retributivism is not applicable in transitional circumstances because punishment cannot fully vindicate the victims of past wrongdoings (she has not specifically targeted expressivist retributivism when she rejects the applicability of retributivism). Her primary reason is that wrongdoings before transitions are normalised and institutionally endorsed: They are normalised because they are so widespread that they become 'a basic fact of life for individuals' (55); they are institutionally endorsed in the sense that 'some citizens or group of citizens have a lower institutionally recognized status in the political community' (93). Murphy points out that victims' diminishment in this context is due to the message implicit in the immediate wrongdoings they encounter and the community's view that they are inferior to others. From here, she contends that 'isolated punishment of a single individual is not going to be effective in achieving its putative retributive purpose' (94). It is because isolated punishment not only 'leaves untouched the actions of others complicit in a particular wrong' (94) but is also 'not likely to be powerful enough to counter the broader institutional context in which unequal treatment was and often still is sanctioned and justified' (94). Then, since isolated punishment cannot 'fully address the source of the challenge to the victim's status' (94), she concludes that retributive justice is not a proper goal of transitional justice.

Defending vindicative retributivism

The above challenge has two problems. First, it is unreasonable for Murphy to assume that punishment in transitional circumstances must take the form of isolated punishment against single perpetrators. In both non-transitional and transitional contexts, institutions of punishment can try and punish complicit wrongdoers. In the UK, a secondary party that assists or encourages a principal perpetrator to commit an offence is also subject to punishment (The Crown Prosecution Service, 2019). Besides, according to Wittmann (2003), during the Frankfurt Auschwitz Trial in the 1960s, 20 officers of the Auschwitz concentration camp were convicted of violating the 1871 German penal code. Seven defendants were convicted of murder, and another 13 were convicted of aiding and abetting murder – the court found them guilty based on their complicit involvement in the murders. These examples show that punishment in transitional circumstances does not necessarily leave out complicit wrongdoers. Criminal justice systems can address complicit wrongdoers to vindicate their diminishment of victims' value. Although isolated punishment would likely be not powerful enough to counter the broader institutional context, the mass punishment of all principal and complicit wrongdoers may do so by sending a much stronger message of moral equality.

Second, it is unreasonable for Murphy to assume that vindicative retributivism justifies punishment only if it can *fully* vindicate the victims' value. For Hampton (1992), partial vindication is still valuable and can justify punishment. She summarises her account of retributivism with a parallel between retributive justice and corrective justice. As she writes, 'Corrective justice compensates victims for harms, whereas retributive justice compensates victims for moral injuries' (1698). Regarding corrective justice, we usually think that compensation is valuable even if it cannot fully repair the harms suffered by the victims. Take the example of a state compensating the family members of a soldier who has died during a military exercise. Although no sum of money can fully compensate for the soldier's death,

we still think some monetary compensation from the state is (morally) valuable. Hampton's parallel suggests that partial vindication, like partial compensation for harms, is valuable even if it fails to repair victims' diminishment fully. For example, in a misogynistic culture, punishing perpetrators of sexual violence alone is not likely to fully re-establish the victims' status. Even if punishment cannot provide sufficient counterevidence against the inferior status of women, some such evidence is still better than none. Even if mass punishment is not powerful enough to counter the diminishing view of political institutions, it can still be valuable to the extent it partially vindicates the victims' diminishment.

For the above two reasons, I conclude that Murphy is wrong to claim that punishment cannot achieve vindicative goals in transitional circumstances. Even if we grant Murphy's implicit assumption that vindication is the fundamental goal of retributive punishment, she has still failed to show that retributivism is not applicable in transitional circumstances.

Defending expressivist retributivism

Lastly, I will also examine whether Murphy's arguments defeat the applicability of expressivist retributivism in transitional circumstances. Within the framework of expressivist retributivism, the goal of punishment is to give culpable wrongdoers what they deserve. Can this goal be achieved in transitional circumstances? Let's consider Murphy's (2017) claim that isolated punishment of single perpetrators 'leaves untouched the actions of others complicit in a particular wrong' (94). Complicit agents of wrongdoings can also be culpable, so if punishing complicit wrongdoers was impossible in transitional circumstances, the goal of expressivist retributivism could not be met. However, I have shown that criminal justice institutions can try and punish complicit wrongdoers in both non-transitional and transitional circumstances. This possibility resolves Murphy's potential challenge to the applicability of expressivist retributivism.

Murphy also argues that punishment in transitional circumstances cannot address the political institutions that treat some citizens as having inferior status. Granting the truth of the claim, it still does not show that expressivist retributivism is inapplicable in transitional circumstances. Even if punishment cannot give political institutions what they deserve, retributive justice can still be achieved by giving individual wrongdoers what they deserve – retributive desert claims arise on a case-by-case basis, and each of their satisfaction still has intrinsic value.

Furthermore, contrary to Murphy, I also argue that punishment in transitional circumstances *can* address discriminative political institutions. It is because institutional moral agents can also be punished.[1] Collective punishment may, at first sight, sound confusing because groups and institutions have, in Erskine's (2011) words, 'no soul to be damned and no body to be kicked' (277). Many historical instances of 'collective punishment', which often came as wars, also run into the problem of guilt by association – punishing individuals for their associations rather than their actions. Despite these difficulties, Erskine argues that it is possible to punish institutional agents through 'dismantlement, boycott, and "naming and shaming' (286). Rather than misdirecting institutional punishment to innocent individuals, these methods

[1] I would not defend the ontology of institutional moral agents here because readers who disagree with this assumption will find no need for punishment to address discriminative political institutions within the framework of expressivist retributivism.

punish institutions *as institutions*. Similarly, Vernon (2011) suggests that a political collective can be punished 'in its political aspect' (304) in ways including 'incorporating it within a larger state' (304), 'dividing it into several pieces' (304), 'imposing regime change upon it' (304), and 'imposing strict limits on a state's political capacity, particularly its military capacity' (304), such as how the Treaty of Versailles limited German rearmament and Japan's disarmament after World War II.

Thus, contrary to Murphy, I conclude that punishing atrocity perpetrators in transitional circumstances can achieve retributive goals, regardless of whether they are defined by expressivist or vindicative retributivism. We now have one less reason to think that retributivism is an inapplicable justification of punishment in transitional circumstances.

Challenge Two: No Authority Has the Standing to Punish

In this section, I will address Murphy's another challenge that no authority has the standing to punish atrocity perpetrators in transitional circumstances. I will first explain why the existence of some authorities with the standing to punish atrocity perpetrators is necessary for punishment to be justified. I will then introduce Murphy's challenge and Lenta's response to it. Lastly, I will argue that Lenta's response is insufficient and offer my response to the challenge.

The standing to punish

Although (expressivist) retributivism holds that culpable wrongdoers deserve punishment, it does not imply that anyone can punish the wrongdoers. Justified punishment also requires that the punisher has the standing to inflict punishment. As Duff (2001, 185-188) argues, an agent loses the standing to accuse or punish someone for flouting certain values when she has shown no respect for those values. His example is that an academic institution that indulges its staff in lateness lacks the standing to discipline students for late attendance. Todd (2019) similarly contends that 'one has standing if and only if one is morally committed to the values that condemn the wrongdoer's actions' (357). If an agent blames a wrongdoer but lacks the standing to blame, the wrongdoer may justifiably reply, 'who are you to blame me for this anyway? You are in no position to blame me' (349). Thus, for the punishment of atrocity perpetrators to be justified, some authorities must be committed to the values that the perpetrators have flouted; otherwise, punishing them is not justified even if they deserve punishment.[2]

Murphy's challenge

Murphy's (2017) challenge is that the above precondition of punishment is not met in transitional circumstances. She begins by pointing out that transitional circumstances are characterised by 'pervasive structural inequality' (43). Structural inequality occurs when the basic structure of society – the institutions that 'regulate and shape interaction among citizens and officials in political societies' (44) and thus have a pervasive influence over

[2] I would not investigate the possibility that some individuals have the standing to carry out 'private justice' due to the widely known dangers of vigilante. Even if it is true that some individuals have the standing to punish and can avoid the dangers of vigilante, their 'private justice' is not needed because, as I will show very shortly, there exist authorities that have the standing to punish atrocity perpetrators in transitional circumstances.

citizens' life prospects and opportunities – constrains and enables similarly situated citizens in different ways that lead to unfair outcomes. And '[s]tructural inequality is pervasive when many different kinds of institutions are defined by terms that are unequal (the horizontal dimension), and the inequality permeates each institution (the vertical dimension)' (47). She argues that structural inequality in transitional circumstances is so pervasive that 'it passes a threshold beyond which the legitimacy of the institutional order is in doubt in the sense that people have reason to adopt certain kind of measures to overthrow that order' (49).

This feature of pervasive structural inequality leads Murphy to argue that the transitional state lacks the standing to inflict punishment. The pervasive structural inequality implies that the state is not committed to 'the equal status of all citizens, which wrongdoing challenges' (93). Moreover, if the state was implicated in the past atrocities, 'it is not neutral with respect to wrongdoing as the retributive account assumes' (93). She thus concludes that the transitional state is 'not necessarily in the best position to ensure that the appropriate punishment is given' (93).

Lenta's response and why it fails

In response to Murphy, Lenta (2019) argues that a transitional state can regain its standing to punish by taking two actions. First, 'it acknowledges its involvement in and sincerely accepts its share of the responsibility for human rights atrocities' (391). Second, 'it admits its past facilitation of, sincerely accepts its share of responsibility for, and rectifies the pervasive structural inequality it previously supported' (391). After the transitional state has taken both actions, Lenta suggests that it can 'achieve sufficient standing to punish perpetrators of gross human rights violations occurring during the pre-transitional conflict' (391).

I agree with Lenta's idea that agents that have lost their standing to punish can nonetheless regain their standing by recommitting to the relevant values. Suppose there is a parent who regularly started bar fights when she was young. I think it is intuitive that she can regain the standing to discipline her child for hurting others if she has stopped engaging in fights and recommitted to peace. I also agree that if a transitional state has completed the two tasks suggested by Lenta, it will regain the standing to punish atrocity perpetrators.

However, I argue that Lenta's response still fails to show that retributivism is applicable *in transitional circumstances*. Once the transitional state has completed both tasks, it is no longer transitional – it has done well enough to 'graduate' from transitional circumstances. I agree with Lenta that the above two strategies are sufficient to let a transitional state regain its standing to punish. However, they fail as a response to show that retributivism is applicable *in transitional circumstances*. Murphy assumes that pervasive structural inequality is a defining feature of transitional circumstances. Since Lenta does not contest this characterisation, I assume he accepts this point. But this is precisely why his response fails: His second strategy for regaining the standing to punish requires the state to rectify pervasive structural inequality; however, once the state is no longer characterised by pervasive structural inequality, it will have already moved out of transitional circumstances. In other words, what Lenta has suggested is paradoxical: To regain the standing to punish in transitional circumstances, the state needs to move out of transitional circumstances. He has not shown how a state can regain the standing to punish despite the unique challenges of transitional circumstances. He has merely solved Murphy's challenge by assuming it away.

Therefore, his response fails to show that retributivism is applicable in transitional circumstances.

Addressing Murphy's challenge

In the following, I will argue that there do exist authorities with the standing to punish atrocity perpetrators in transitional circumstances. Even if Murphy is right that transitional states cannot have the standing to punish atrocity perpetrators, there are still other authorities that have the standing to do this. As Hansen (2017) describes, 'transitional justice increasingly occurs in a variety of spaces, including the local and international, which were seen as less relevant in the early field' (51). States are not the only potential agent who can punish atrocities; international agents also have this capacity. The Rome Statute of the ICC entered into force on 1 July 2002, and it currently has jurisdiction in 123 member states to punish international crimes, including genocide and crimes against humanity (ICC, 2022). In the Democratic Republic of Congo, Uganda, the Central African Republic, and Mali, states have made referrals to the ICC to undertake prosecutions for atrocities on their territories (Kalyk, 2018). To try the agents of atrocities in former Yugoslavia and Rwanda, the UN also established ad hoc international tribunals, including the ICTY and the ICTR (United Nations Security Council, 2021). Furthermore, in Cambodia, East Timor and Sierra Leone, the domestic states have cooperated with the UN to create hybrid courts to try past international crimes (Linton, 2001).

The courts and prosecuting agencies in the above cases are composed of purely international personnel or local and international personnel. These international agencies have the standing to punish atrocities for two reasons. First, while transitional states might be implicated in the past atrocities, the international agencies would likely be not involved, so there would likely be no evidence against their commitment to the values flouted by the atrocity perpetrators. Second, these international agencies are often chosen to represent the shared moral values of the international community. Therefore, as Linton (2001) argues, international involvement 'brings in much-needed objective expertise and minimises the risks of partiality and lack of independence that arise from local participation in the process' (186). Thus, regardless of whether transitional states have the standing to punish, there are still international authorities that satisfy the precondition of retributive justice. Murphy's claim that retributive justice cannot be pursued in transitional circumstances due to the lack of proper punisher is invalid.

Here, a critic (or Murphy) may challenge whether international organisations and foreign states have the jurisdiction – the right to use coercion – to punish individuals in the transitional state. Normally, we assume that states have sovereignty right over their territory; unless they voluntarily invite the involvement of international agencies, outsiders have no right to prosecute or punish individuals within their territory. If this assumption holds in transitional circumstances, international agencies will lack the jurisdiction to subject atrocity perpetrators to coercive punishment even though they have the moral standing to do so.

The assumption that transitional states retain full sovereignty right over their territory is, however, precisely defeated by Murphy's (2017) characterisation of transitional circumstances: The structural inequality in transitional circumstances is so pervasive that 'it passes a threshold beyond which the legitimacy of the institutional order is in doubt in the

sense that people have reason to adopt certain kind of measures to overthrow that order' (49). The conditions of state legitimacy – the conditions under which a state can claim the right of territorial jurisdiction – draw the boundaries between the existence and absence of not only the citizens' right to resist but also outsiders' right to intervene. Given the citizens' right to self-determination, political institutions that represent the citizens should eventually fill the gap in territorial jurisdiction. But during the transition window, when the self-determining institutions are still building up their capacity and demonstrating their value commitments to gain full legitimacy, outsiders have the duty to help the transitional state provide justice within the territory and the corresponding right to achieve justice through coercive means. In other words, this challenge of pervasive structural inequality identified by Murphy is, at the same time, an 'opportunity' that gives international agencies the right to inflict punishment on atrocity perpetrators in the transitional state's territory.

To conclude, although I agree with Lenta that transitional states can take actions to regain the standing to punish, this response does not show that transitional states can have the standing to punish before the completion of the transition. The ultimate reason for the failure of Murphy's challenge is this: She has ignored that international agencies can also have the standing and jurisdiction to punish atrocity perpetrators during transitional circumstances. There is, hence, another fewer reason to think that retributivism is an inapplicable justification of punishment in transitional circumstances.

Challenge Three: Punishment Is Ex Post Facto

Murphy's next challenge is that punishing atrocity perpetrators in transitional circumstances is ex post facto. I will first introduce her challenge. I will then explain Lenta's response and address potential objections to it.

Murphy's challenge

Murphy (2017) begins her challenge by pointing out that 'punishment for normalized wrongdoing will be, at least in some cases, ex post facto' (91). For her, ex post facto punishment punishes individuals for actions that did not violate any legal norm that was in effect at the time of action. She observes that 'many morally egregious acts may not have been legally proscribed by declared rules at the time they occurred' (92); and even if they are formally prohibited, they may not have been enforced regularly. Based on this observation, she suggests two reasons for why retributive justice cannot guide the pursuit of transitional justice. The first reason is this: 'A basic premise of all theories of criminal punishment, both the utilitarian and retributive kind, is that punishment should not be ex post facto' (91). If retributivism presumes that punishment should not be ex post facto, it will not justify ex post facto punishment. The second reason is this: 'Ex post facto punishment violates the commitment to prospectivity at the core of the rule of law' (91). This reason appeals to the principle of legality, a core element of the rule of law. Together, these two reasons show that '[a]ny justification of punishment in the circumstances of transitional justice, then, must explain why ex post facto retribution could be permissible and count as retribution' (92). If this cannot be shown, retributive justice will be inapplicable in transitional circumstances.

Retributivism does not presume non-retroactivity

In response to the first reason, Lenta (2019) argues that retributivism does not presume that punishment is not ex post facto. He argues that the desert bases of punishment are not criminal but moral wrongdoings. Thus, even if torture, rape, and murder have been lawful at the time of commitment, 'something of intrinsic value is realized' (389) when the perpetrators receive the punishment they deserve. This line of reasoning also applies to expressivist retributivism, according to which proportional punishment should be inflicted to match wrongdoers' moral rather than legal culpability. Ex post facto punishment that can still give culpable wrongdoers what they deserve and thus achieve retributive justice. Therefore, Murphy's argument that retributivism presumes the non-retroactivity of punishment is invalid.

The value of legality

In response to the second reason, Lenta (2019) acknowledges that the principle of legality 'obliges the state to give people fair warning' (388). However, he denies that legality constitutes an absolute moral reason against retroactive punishment; for him, 'there could be circumstances in which legality-related reasons against punishment of the perpetrators in question are outweighed by countervailing reasons in favour of punishment' (388). As Lenta continues, these countervailing reasons are predominantly the great intrinsic value of giving the perpetrators of serious wrongdoers what they deserve. He also finds support for this 'lesser-evil' view from H.L.A. Hart, Jeffrie Murphy, Leo Zaibert, and Michael Moore. As such, even if Murphy is right that ex post facto punishment violates the commitment to legality, this reason is not necessarily strong enough to defeat all retributive desert claims.

Although I agree with Lenta's conclusion that retributive justice can outweigh the value of legality, I think his argument is not strong enough to persuade people who hold dearly to the principle of legality. The judgment of the relative weight between retributive justice and legality will, ultimately, depends on what significance we assign to each of them. I think it would be helpful to strengthen Lenta's argument by attending to the other side of the equation – there are good reasons to think that the weight of legality-related reasons against punishing atrocity perpetrators is relatively weak.

Why should punishment be non-retroactive? The legal literature commonly recognises that legality protects individual rights and liberties from power abuse (Shahabuddeen, 2004; Mokhtar, 2005; Dana, 2008; Gallant, 2009; Galand, 2019). This protection is needed because punishment can place extensive limitations on individual rights and liberties. If the state abuses its coercive power through arbitrary detention and punishment, citizens' fundamental interests will be seriously harmed. In Gallant's (2009) words, 'the criminal law applies the highest legal sanctions available to a society: deprivation of freedom, confiscation of property, and in some societies, death' (17). Therefore, any criminal law system 'has the strength, if misused, to commit abuses' (405). Here, the principle of legality is valuable because, by obliging the state to provide fair notice to citizens about the definitions of crimes and their respective penalties in advance of inflicting punishment, it protects individual citizens from power abuse: With fair notice, citizens can foresee the legal consequences of their actions and adjust their behaviour to avoid criminal liability, and hence avoid being deprived of property, freedom, and/or life. Even if the law announced is unjust, citizens can still choose to comply with it to avoid further deprivation of freedom. Most importantly, the principle of legality prohibits the state from punishing dissenters and its

opponents on a whim. Legality restricts the state's scope of justified coercion by obliging it to rely on previously enacted laws.

The strength of legality-related reasons

There are two reasons why legality does not constitute a strong reason against punishing atrocity perpetrators in transitional circumstances. First, because culpable wrongdoers do not possess moral rights that stand in the way of proportional punishment, the protection of legality does not extend to them. When a person commits an act of wrongdoing, she forfeits her rights against the imposition of hard treatment. As Morris (1991) argues, rational humans are owed full respect as moral agents only if they are 'willing to impose constraints [of justice] on their behaviour' (62), so if they have acted in ways that violated these constraints, they would 'lose some of their rights and some of their moral standing' (62). As such, when we impose hard treatment on someone who has forfeited her rights, 'we commit no injustice against this individual for he or she lacks the relevant moral rights that would otherwise make such treatment unjust' (68). The value of legality is grounded in protecting individual rights, but for wrongdoers who have partially forfeited their rights, protecting that forfeited part is of no value.

I do not mean that legality is a redundant principle: Retroactive punishment tends to be abused and inflicted on innocent citizens, and the moral rights of these innocent citizens do need protection. I only mean that if the targets of retroactive punishment are really culpable for wrongdoings, the protection of legality does not apply to them, so the violation of legality, in this case, is insignificant. Therefore, suppose the 'atrocity perpetrators' punished in transitional circumstances are really culpable perpetrators of atrocities (but not victims of a miscarriage of justice or political prosecution), the principle of legality does not constitute a moral reason against punishing them retroactively.

Here, a critic may argue that ex post facto punishment still carries a significant risk of impermissibly depriving the rights of innocent citizens. And unless we can be entirely certain that the targets of retroactive punishment *are* culpable wrongdoers, the principle of legality still constitutes a reason against retroactive punishment due to this risk. I agree with this analysis, but I do not think that it makes the reason of legality strong enough to outweigh retributive justice in all cases of punishing atrocity perpetrators. It is because this risk can be mitigated by the involvement of impartial actors with solid capacities for collective evidence. In the previous section, I agreed with Murphy that transitional states lack the standing to punish atrocity perpetrators and suggested that this could instead be done by international agencies with the standing and jurisdiction necessary for inflicting punishment. Again, as Linton (2001) observes, the involvement of international actors 'brings in much-needed objective expertise and minimises the risks of partiality and lack of independence that arise from local participation in the process' (186). When states inflict retroactive punishment on their citizens, we think legality hugely matters because the punishers are often embroiled in local political conflicts, and so there is a significant risk that they are abusing their power to harm their innocent opponents. In contrast, the more distant position of international actors minimises this risk and makes the strength of the reason of legality relatively weak. This is another benefit that the unique 'opportunity' of transitional circumstances offers. Therefore, I insist that legality does not constitute a strong reason against punishing atrocity perpetrators in transitional circumstances.

The second reason is that not all cases of punishing atrocity perpetrators will be ex post facto. In states that have ratified the Rome Statute of the ICC, citizens have received fair notice of the criminality of atrocities. Punishing atrocity perpetrators is thus not ex post facto for atrocities that occur after the state ratifies the Statute. There are no legality-related reasons against the punishment of atrocity perpetrators in these cases. Furthermore, in states where no (national or international) law criminalises atrocity, they would still likely have criminalised many elements of atrocities, such as murder, assault, enslavement, sexual violence, false imprisonment, etc. Lenta (2019), for instance, observes that in apartheid South Africa, state agents were well aware of the illegality of their human rights abuses, so their actions were 'typically carried out covertly, to avoid detection, prosecution and punishment' (389). In these cases, punishing atrocity perpetrators is also not ex post facto and therefore raises no legality concerns.

Some may worry that punishing atrocity perpetrators *for committing atrocities* is not fair because they have only received notice that they would be punished *for ordinary crimes*. The worry is that the punishers would abuse the target's guilt for the ordinary crime as an excuse to inflict whatever kind and level of punishment they wish. This possibility of power abuse then invites the protection of legality. However, as Galand (2019) shows, the reclassification of a national crime into an international crime is consistent with the principle of legality if 'the crime must have existed under domestic applicable law at the time of commission and the sentence cannot be higher than the one provided for by the domestic applicable law' (149). Reviewing the judgments of international courts, both Gallant (2009) and Galand (2019) find that the courts have commonly held the view that ordinary crimes can be reclassified as atrocity crimes without violating legality. For instance, according to Galand (2019), the ICC can, without violating legality, punish individuals submitted from states not party to the Rome Statute by looking at 'the domestic law applicable to the conduct to see whether a crime sufficiently similar to the Rome Statute's provision' (148), and 'if the Court finds that the underlying acts were criminalized under applicable national law, the offence can also be reclassified as an international crime' (148).

Another example is a case ruled in Spain involving Adolfo Scilingo, a former navy captain in Argentina (Kalyk, 2018). Article 607bis of the Spanish Penal Code entered into force in 2004 and criminalised crimes against humanity. In 2007, the Spanish Supreme Court confirmed that Scilingo was guilty of crimes against humanity due to his involvement in throwing government dissenters out of aeroplanes during Argentina's military rule from 1976 to 1983. Although Article 607b is extraterritorial and the acts were committed before the article entered into force, the Supreme Court held that 'the conviction did not violate the principle of legality given the crime was clearly foreseeable at the time of the offence as the underlying crimes of murder and unlawful detention already constituted crimes under domestic law' (458). I am not claiming that all cases of punishing atrocity perpetrators are not ex post facto; I only mean that many of these cases will not violate the principle of legality when the conditions of reclassification are met. To the extent that the elements of atrocities are already criminal under the national penal code, the strength of legality-related reasons against punishing atrocity perpetrators will be weakened.

A critic may also point out that even if national laws prohibit the elements of atrocities, they may not be regularly enforced: In this kind of cases, it is reasonable for the criminals of ordinary crimes to expect no punishment, and so punishing them still violates the principle of legality. As Murphy (2017) writes, 'In cases where the legal rules are not generally enforced, the same underlying moral concern with ex post facto prosecution is salient' (92). It is because enforcing previously unenforced rules is to 'hold citizens and officials to a standard of conduct they did not reasonably believe applied at the time of action' (92).

However, I argue that enforcing previously unenforced rules does not violate the principle of legality. Murphy's argument assumes that punishment complies with legality only if it is reasonable for the target to believe that the law will be enforced at the time of the violation. This assumption is problematic for two reasons. First, this assumption is inconsistent with how legality is conventionally defined. The legal literature (Shahabuddeen, 2004; Mokhtar, 2005; Dana, 2008; Gallant, 2009; Galand, 2019) defines legality as requiring the state to give fair notice of the criminality of actions but not to make citizens reasonably believe that the corresponding law will be enforced. Second, if this assumption is true, legality will facilitate rather than prevent power abuse. Suppose some officials think that a particular law limits their power to do what they want. Then, by choosing not to enforce that law when it is violated, the officials can make citizens reasonably believe that the law will not be enforced. Because, under Murphy's assumption, legality requires that it must be reasonable for the criminal to believe that the law will be enforced, the officials have successfully created immunity for violating that law. In other words, Murphy's assumption does not prevent power abuse but instead puts officials above the rule of law by giving them the power to create immunity for themselves. For the above two reasons, I reject Murphy's assumption and conclude that enforcing previously unenforced rules does not violate the principle of legality.

To conclude, although I reject Murphy's claim that retributivism presumes the non-retroactivity of punishment, I acknowledge her argument that ex post facto punishment violates the principle of legality. However, following Lenta, I argue that the value of legality does not necessarily outweigh the achievement of retributive justice in punishing atrocity perpetrators. It is because the moral seriousness of atrocities gives rise to strong retributive claims; this is also because legality-related reasons against punishing atrocity perpetrators are relatively weak in transitional circumstances. Their weakness is partly due to the forfeiture of rights of culpable wrongdoers and partly due to the fact that atrocities or elements of atrocities are often already prohibited by national and/or international laws applicable at the time of committing atrocities.

Conclusion

In this chapter, I have investigated the thesis sub-question of whether retributivism is an application justification of punishment in transitional circumstances. Having outlined the expressivist theory of retributivism, I then addressed three of Murphy's four challenges against the applicability of punishment. Regarding the challenge that punishment cannot achieve retributive goals in transitional circumstances, I have argued that Murphy wrongly assumes that punishment can only target individual perpetrators but not complicit wrongdoers; I have also shown that collective punishment allows discriminative institutional

agents to be addressed. Regarding the challenge that no authorities have the standing to punish atrocity perpetrators in transitional circumstances, I have argued that the transitional state's lack of full legitimacy allows international agencies with the standing to punish to punish atrocity perpetrators. Regarding the challenge that punishing atrocity perpetrators is ex post facto, I have argued that the value of legality can be outweighed because legality-related reasons against ex post facto punishment are relatively weak in transitional circumstances. Therefore, I conclude that expressivist retributivism is an application justification of punishment in transitional circumstances. Since expressivist retributivism treats culpability as a necessary condition of justified punishment, the principal thesis question of whether civil servants should be punished for past state atrocities then depends crucially on whether civil servants can be held repsonsible for atrocities. This will be examined in the subsequent chapters.

Chapter 3
Complicit Responsibility

In the previous chapter, I argued that retributivism offers an applicable justification for punishing past atrocities in transitional circumstances. I have outlined the expressivist theory of retributivism, according to which punishing an individual is permissible and justifiable only if she is blameworthy and thus deserves punishment.

Next, this chapter will deal with the following sub-question: What are the conditions of holding civil servants complicitly responsible for past state atrocities? To answer this question, I will investigate the minimal conditions of holding individuals responsible for collective wrongdoings. In other words, what are the minimal criteria of complicit responsibility? This investigation is the next step in tackling the thesis question of whether civil servants should be punished for past state atrocities in transitional circumstances. The first step has been taken in the previous chapter to defend the applicability of retributivism in transitional circumstances. Given that culpability is a necessary condition of justifiable punishment, and that atrocities are serious wrongdoings, civil servants should be punished for atrocities only if they are morally responsible for the atrocities. So, are they? This is the guiding question of this and the next chapter. This chapter will examine the minimal criteria for holding them responsible; the next chapter will then inspect some of their common defences to determine whether at least some of them meet those minimal conditions.

This chapter will also address Murphy's (2017) last challenge to retributivism's applicability in transitional circumstances. In this challenge, Murphy first shows that retributivism requires that punishment is proportional to individual responsibility. She then claims that 'theories of retributive justice do not provide resources for the appropriate way to determine individual responsibility for collective crimes' (89). It is because individuating responsibility for collective wrongdoings requires giving up the assumption that individual responsibility should be assessed based on 'what an individual caused on his or her own, consequences that are assumed to be under his or her control' (90). As such, she argues that this difficulty in satisfying the proportionality requirement makes retributivism inapplicable in transitional circumstances. This chapter will address this challenge by rejecting the above assumption.

This chapter will proceed as follows. I will first conceptualise complicit responsibility and introduce the conditions of individual responsibility. Next, I will introduce the problem of overdetermination and the different criteria of complicit responsibility that the literature proposes to overcome the problem. Lastly, I will reject these proposals and argue that individuals can be held responsible for actions that carry the risk of making collective wrongdoings more likely to succeed. Apart from setting the stage for determining civil servants' responsibility for past atrocities, this chapter also aims to advance contemporary debates on complicit responsibility.

Defining Complicit Responsibility

Complicit responsibility is individual responsibility for collective wrongdoings. As does individual responsibility for individual wrongdoings, complicit responsibility also calls

an individual moral agent into account (in contrast, collective responsibility for collective wrongdoings calls collective agents into account). However, complicit responsibility is also different because it assesses individual actions based on their connections with collective wrongdoings – wrongdoings of which the primary agents of action are collectives – whereas individual responsibility for individual wrongdoings is only concerned with wrongdoings of which the primary agents of action are individuals.

There are three crucial things to note about this definition of complicit responsibility. First, it assumes that there are genuinely collective actions that should primarily be ascribed to collective agents instead of individuals. This assumption underlies the current practices of international law, which 'attributes holistic responsibility to states for their acts' (Stilz, 2011, 190). Following this assumption, atrocities are not the mere aggregation of individual wrongdoings, but a kind of collective actions executed by collective agents in a 'planned, systematized, and organized' (Drumbl, 2007, 4) manner.

Second, some may reject the idea of complicit responsibility because they believe that collective responsibility excludes individual responsibility. This position, however, contradicts the conviction that all actions of individual moral agents are subject to moral assessment. Given that individual actions are a necessary component of collective wrongdoings, ruling out individual moral responsibility would entail not subjecting any of these individual actions to moral assessment. Thus, collective responsibility co-exists with individual responsibility.

Third, some may reject defining complicit responsibility as 'individual responsibility for collective wrongdoings' because they believe that an individual can only be responsible for her own actions. For instance, Isaacs (2011) argues that 'individuals cannot be responsible for collective actions as such' (97), but only for 'their contributions to collective actions' (97). The underlying worry here is that holding individuals responsible for collective wrongdoings would result in 'guilt by association', which means that 'a degree of solidarity within a group, or a shared aspect of identity, allows those who are not party to a specific action to be morally praised or blamed for the action of an agent or agents within the same group' (Erskine, 2011, 275). Essentially, guilt by association 'runs counter to the understanding that one cannot be blamed for the wrongdoing of another' (276).

My conception of complicit responsibility does not entail guilt by association, for this conception is also built on the premise that an individual can only be responsible for her own actions. My investigation presumes that if an individual is responsible for a collective wrongdoing, the basis of her responsibility must be some connection between her action and the collective wrongdoing, but not any emotional attachment or shared identity that she has with the collective. Solidarity may correlate strongly with complicit actions, but it cannot be itself the basis of complicit responsibility. As long as the object of moral assessment is individual actions, the difference between an individual being 'responsible for a collective wrongdoing' and 'responsible for her contribution to a collective wrongdoing' is merely linguistic.

The Criteria of Individual Responsibility

Now, because complicit responsibility is a sub-category of individual responsibility, it is best to begin the investigation by understanding the minimal criteria of individual responsibility. By examining the practice of holding each other responsible, and in particular, how responsibility can be excused or exempted, Wallace (1994) argues that a person is blameworthy for her action if it is unjustified and expresses a culpable choice – a choice that reflects she has failed to grasp and apply moral reasons which she should have grasped and applied (120-121, 128 & 226). It is because the practice of holding people to moral obligations involves a commitment to the possibility of justifying the obligations with 'reasons that could move both the judge and the agent held responsible to comply with the obligations in question' (130). And since choices are 'susceptible to being influenced directly by reasons' (131), the commitment to justification can be sustained only if the assessment of compliance or violation of moral obligations targets the quality of choices. Furthermore, because the commitment to justification also presumes that reasons can move the person, Wallace argues that a person must possess the 'powers of reflective self-control' (155) to be held responsible as a moral agent. The powers are 'the general ability to grasp and apply moral reasons and to regulate their behavior by the light of such reasons' (155).

The advantage of Wallace's account of individual responsibility is that it justifies three essential aspects of our daily moral practices. First, treating culpable choice as a condition of blameworthiness explains 'why we should attach so much moral importance to the question of whether the bodily movement that a person has made was or was not intentional' (133). Second, this also explains how excuses mitigate or remove an agent's responsibility for particular actions: Excuses function not by denying that the agent has taken the action, or the action was unjustified; instead, excuses show that action does not express a culpable choice, so the agent has not really violated any obligations. One of Wallace's examples is that if another person forcibly uses one's hand to beat a third party, the hand's owner will not be responsible for beating the third party, for the 'beating' is really an unintentional bodily movement that does not express any culpable choice of the hand's owner. Third, the account explains how exemptions mitigate or remove a person's responsibility for all actions. It is because people with underdeveloped or impaired practical reasoning and self-control capacities do not fully possess the powers of reflective self-control, and therefore do not fully meet the conditions of being an individual moral agent.

Wallace's account provides a clear guide for assessing individual culpability in the context of individual wrongdoings. Having learnt that an individual's action is unjustified, we then ask whether she qualifies as a moral agent and whether her action expresses a culpable choice. However, in the context of collective wrongdoings, it is only given that the collective action is unjustified. It is not clear which individual actions, or whose actions, are unjustified. Hence, to understand the minimal criteria of complicit responsibility, we need to understand what kinds of connections between individual actions and collective wrongdoings are relevant to our moral assessment. In other words, we look for connections that explain how individual actions can inherit the moral quality of collective wrongdoings.

The Problem of Overdetermination

A commonsense answer to how an individual becomes responsible for a collective wrongdoing is that her action causes the collective harm. For instance, a general who

ordered an army to execute a genocidal programme should be held responsible for it, for, without her order, the targeted group would not have been destroyed. However, this answer is considered inadequate in the literature due to the problem of overdetermination. As such, other ways of holding individuals responsible for collective wrongdoings have been proposed. I will introduce the problem of overdetermination in this section and review the solutions to this problem in the next.

The problem of overdetermination arises from four fundamental premises. The first premise describes a necessary condition for holding an individual responsible for a harm:

(P1) An individual can be held morally responsible for a harm only if her action has caused the harm.

Both Kutz (2000) and Lawson (2013) identify P1 as a core element of our commonsense assessment of individual responsibility. As Kutz (2000) writes, 'Paradigmatically, individual moral agents are reproached, or reproach themselves, for harms ascribable to them and them alone, on the basis of their intentional actions and causal contributions' (4). Similarly, Lawson (2013) states that 'we are accustomed to regarding individuals as responsible for outcomes to which their actions make a causal difference and over which they have causal control' (230).

The second premise defines causation as counterfactual dependence:

(P2) An action causes a harm if and only if it has 'made a difference' – if the action had not occurred, the victim would not have been harmed, or would not have been harmed that seriously.

P1 and P2 entail the following conclusion:

(C1) An individual can be held morally responsible for a harm only if her action has made a difference to the harm.

Kutz (2000) calls C1 the 'Individual Difference Principle': 'I am only accountable for a harm if something I did made a difference to its occurrence. If substantially the same harm would have occurred regardless of what I have done, I cannot be accountable for it' (3). And he suggests that it is a fundamental principle in our 'common, nonphilosophical understanding of individual accountability' (3).

The third premise, then, introduces the case of overdetermination:

(P3) In any case of fully overdetermined collective harm, no individual action has made a difference to it.

This means that, for each individual present in a case of fully overdetermined collective harm, even if the individual had not done what she has done, the collective harm would still have been as serious as it is. This can also be expressed in possible worlds semantics: In the closest (most similar to the actual world) possible world among all where she had acted differently, the collective harm would still have been as serious as it is in the actual world. I will illustrate the possibility of fully overdetermined

collective harm with two examples. The first example is from Lepora and Goodin (2013): Four marksmen 'fire and their bullets strike the victim's heart simultaneously, resulting in his death' (64). Had any marksman not fired, the victim would still have been killed. Thus, no single marksman has made a difference in this outcome. The second example is the Allied aerial bombing of Dresden from 13 to 15 February 1945, a historical case widely discussed in the literature of complicit responsibility. According to Kutz (2000), the bombing was estimated to have killed 35,000 civilians. The decimation of the city was the collective outcome of at least 8,000 crewmen performing different roles on the planes. As he notes, 'Each crewman's causal contribution to the conflagration, indeed each plane's, was marginal to the point of insignificance' (118). Even if any of them had refused to drop the bombs, the other planes' explosives and incendiaries would still have destroyed the city.

C1 and P3 lead to the following conclusion:

(C2) In any case of fully overdetermined collective harm, no individual can be held morally responsible for the collective harm.

However, C2 is believed to be inadequate due to the following dearly held premise in the literature:

(P4) In some cases of fully overdetermined collective harm, someone must be held individually responsible.

Some ground this premise in our intuitions. Lawson (2013) suggests that if we only blame individuals for what they cause on their own, 'in some collective contexts there is an intuitive sense in which this is not enough' (228). Similarly, Isaacs (2011) argues that if individuals must have created a causal impact to be held responsible, those with negligible impact 'seem to escape blame even when our intuitions tell us they should not' (126). Kutz (2000) justifies this premise with a more profound metaethical commitment: Theories of individual responsibility need to prevent collective harms from occurring in the first place, so they must 'provide individual agents with reasons to avoid and reasons to respond to even marginal participation in harmful acts' (123). But if no one could be held responsible for any overdetermined collective harm, the system of moral accountability would fail to 'afford us some protection from the serious harms we can so easily bring about together' (7).

The problem of overdetermination is, then, to find a way to account for P4 despite the plausibility of P1, P2, and P3.

Solutions to the Problem of Overdetermination

The literature has proposed different ways to overcome the problem of overdetermination. Each proposal represents an answer to the criteria of complicit responsibility. This section will review Kutz, Lawson, Gardner, and Petersson's solutions.

The theory of inclusive ascription

Both Kutz (2000) and Lawson (2013) suggest supplementing P1 with the theory of inclusive ascription. They agree that individuals can be held responsible for the harm

they cause alone. However, apart from this mode of ascribing responsibility, which they call ‘exclusive ascription’ (Kutz, 2000, 140), there is also ‘inclusive ascription’ (140) of responsibility, where ‘inclusive authorship constitutes a fundamentally different responsive position from exclusive authorship’ (139). Kutz proposes what he calls the ‘Complicity Principle’ (122), according to which a collective harm or wrongdoing can be inclusively ascribed to individuals if they ‘intentionally participate in the wrong they do or harm they cause’ (122). He argues that inclusive ascription is based on the teleological relation between the individual and collective action but not on any causal relation between them. Similarly, Lawson (2013) contends that ‘individuals are responsible for collective wrongs in virtue of a teleological connection to a collective goal without regard for their own causal contributions’ (236).

As such, even if an individual has made no difference to a collective harm – she has not caused it – she can still be held inclusively responsible for her intentional participation in the corresponding collective action. This way, Kutz and Lawson allow individuals to be held responsible in some cases of fully overdetermined collective harm, hence saving P4. For instance, in the aerial bombing of Dresden, since each crew member ‘intentionally does his or her part of realizing a collective goal, raising the firestorm’ (Kutz, 2000, 141), each can be held inclusively responsible for the aerial bombing.

Although the inadequacy of C2 gives plausibility to the theory of inclusive ascription, I reject this theory because it ascribes blame to individuals whom we intuitively think are innocent. By examining the meaning of ‘teleological relation’ and ‘contribution’ in the theory, I will argue that it entails counterintuitive implications in two kinds of cases, namely identity expression and wishful contribution.

Kutz and Lawson repeatedly use the terms ‘contribution’, ‘(intentional) participation’, and ‘playing a part’ to describe their theory of complicit responsibility. We sometimes use these terms to describe individual actions that actually cause collective outcomes. Yet, as shown above, both Kutz and Lawson reject that causal connection is necessary for an action to qualify as contribution; instead, they make teleological relation a sufficient condition for an action to qualify as contribution. As Kutz (2000) writes, ‘What makes my behavior participatory is nothing more (and nothing less) than my conception of what I do as related to the group act’ (82), and ‘claims of inclusive authorship are licensed by the fact that each member’s acts are explained by the overlapping goal of realizing the group’s plans’ (107). (Lawson develops his analysis of Kutz’s theory without redefining ‘contribution’.) In other words, Kutz believes that an individual action qualifies as contribution to a collective action if and only if the individual has a conception of the collective action, and this conception explains her action. This explanatory link obtains in the sense that her action is ‘counterfactually sensitive’ (83) to the collective conception – if she had not had this conception, she would have acted differently. And this explanatory link is what Kutz calls the teleological relation between the individual and collective action. In sum, according to the theory of inclusive ascription, this teleological relation is necessary and sufficient to qualify an individual action as contribution and sufficient to implicate blame for the individual. This implies that an agent’s action can be unjustified because it is counterfactually sensitive to her conception of a collective wrongdoing.

However, the inclusive ascription theory's way of implicating blame will entail counterintuitive implications in two kinds of cases. The first kind of cases is identity expression. To express solidarity with a group, individuals sometimes put on badges, wear certain outfits, and follow specific behavioural rules in life. These practices are not limited to individuals' public life, where their identity expressions are visible to others; they may also do these practices in their private life – ideally, a vegan refuses animal products in both restaurants and kitchens. Visibility and recognition are important but not essential because the practices of identity expression are also crucial to maintaining individual integrity.

According to the theory of inclusive ascription, identity expression qualifies as contribution to collective action. Kutz (2000) identifies four types of contributory relations: causal, constitutive, expressive, and normative. The last two types correspond to what I call identity expression. For Kutz, the relation between an individual action and a collective action is expressive if the action 'exemplifies one's membership in a group or participation in an activity' (82); the relation is normative if the action is counterfactually sensitive to 'norms internal to some group or institution that demand certain behavior (I wear a dark suit as an IBM employee)' (82). Identity expression always qualifies as contribution because only the individual's conception of the collective can explain why she practices identity expression in the first place.

As such, inclusive ascription holds individuals responsible for collective wrongdoings even if they have only expressed their identity with the collective. Suppose an individual wishes to join a radical animal rights organisation after learning about its attacks on farms and animal testing facilities. On the organisation's webpage named 'How to Join Our Mission', she finds a detailed set of dietary rules. Eager to join the group, she immediately adopts the dietary rules in her daily life. Then, one day, she reads on the news that the organisation has set fire to an animal testing facility. The fire killed a researcher and his family on a visit. Is she blameworthy? According to the theory of inclusive ascription, because her dietary changes can only be explained by her conception of the organisation's goals, there is a teleological relation between her dietary changes and what the others do, so her changes qualify as contribution to what they do. In other words, her act of identity expression alone, despite having no causal relation with the arson, is sufficient to make her inclusively responsible for the arson. The theory implies that she is blameworthy, but I think it is intuitive that she is innocent – there is nothing wrong with eating fewer animals. I also would not describe her as having 'contributed', 'intentionally participated', or 'played a part' in the arson. Therefore, I argue that the theory of inclusive ascription over-stretches the scope of blame and implies guilt by association in cases of identity expression.

The next kind of cases where inclusive ascription entails counterintuitive implications is wishful contribution. An individual makes a wishful contribution when she performs an action that cannot cause the success of a collective action, but falsely believes that it can. Lawson (2013) claims that for 'an intentional participant in genocide who, despite her best efforts, fails to contribute to the killings' (241), her failure would not exempt her from being inclusively responsible for others' killings. However, this claim is sound only when her best efforts can help the genocidal programme succeed: Suppose she has transported ammunition to the killing site, but they are not used because the executioners already have enough. Intuitively, her failed effort is still blameworthy. Here, the theory of inclusive ascription does account for an intuition that

ascription by actual causation cannot explain: Because her transportation of ammunition can only be explained by her conception of the genocidal programme, this teleological relation makes her inclusively responsible for the genocide.

However, Lawson's claim loses its intuitive appeal when the individual's best effort is just wishful contribution, meaning there is no possible way it can help the genocide succeed. Suppose, instead of transporting ammunition, she just privately prays for the genocide's success, with a sincere belief that doing so can foster its success. If praying is all she has done, it is intuitive that she is innocent – she should not be blamed for the genocide as it is impossible that praying has anything to do with the genocide's success. Nonetheless, because her conception of the genocidal programme uniquely explains her praying, the theory of inclusive ascription still counts it as contribution to the genocide and entails that she is inclusively responsible for others' killings. Thus, apart from cases of identity expression, the theory of inclusive ascription also over-stretches the scope of blame in cases of wishful contribution. I, therefore, conclude that the theory is not a good solution to the problem of overdetermination.

A necessary element of a set of conditions

Next, I will review Gardner's and Petersson's solutions to the problem of overdetermination. Gardner (2004) disagrees with Kutz's theory of inclusive ascription because it makes the causal connection between an individual action and collective harm unnecessary for holding individuals responsible for the harm. He insists that 'the actus reus for complicity must always, in one way or another, be causal' (829). He thus proposes to save P4 by rejecting P2, the counterfactual dependence view of causation. He argues that an action does not need to make a difference to a harm to cause it; instead, he puts forward the following criterion of causation: One event causes another if the first is 'a necessary element of at least one set of jointly sufficient conditions' (828). Applying this criterion to the aerial bombing of Dresden, Gardner contends that each pilot did cause the city's destruction because each pilot is 'a necessary member of various sets of pilots on the mission who between them ensured Dresden's destruction' (828). If his criterion of causation is correct, individuals can be held responsible for what they cause even in cases of fully overdetermined collective harm.

However, I will argue that Gardner's criterion of causation is inconsistent with our quotidian ascription of causal relationships. Suppose a gang lord ordered two teams of assassins, each with an indispensable role in her team, to assassinate a drug dealer. Team A planted a time bomb in the dealer's bedroom and set the explosion time to 11:00 pm. However, at 10:59 pm, Team B spotted the dealer on her bed, watching TV, and sniped her to death. A few seconds later, Team A's bomb exploded and blew her corpse to pieces. Using Gardner's criterion of causation, since each team's effort was sufficient to kill the drug dealer and each assassin played an indispensable role in the team, every assassin caused the dealer's death. Nonetheless, it is intuitive that Team A did not cause the dealer's death – Team A could be left out entirely in the description of the chain of events leading to the dealer's death. Therefore, although Gardner's criterion of causation can save P4, the price of assigning causal links to causally unrelated events is unacceptable.

Fine-grained view of causation

Petersson (2013) shares Gardner's disagreement with inclusive ascription's abandonment of causality. As Petersson writes, 'Causation is an explicit and essential element in most doctrines of legal liability and it is central in common sense views of moral responsibility' (849). Contrary to Gardner, Petersson accepts the counterfactual dependence view of causation. Instead, he puts forward a 'fine-grained' view of causation. According to this view, changing any minute detail of an event qualifies as making a difference: A person has made no difference only if the event remained exactly the same despite her action. Under this view, 'there are not really any genuine overdetermination cases' (853), for in any case of collective harm, there must be someone whose withdrawal would have changed the exact pattern of the harm, and so her action has made a difference to the harm. For instance, in the aerial bombing of Dresden, '[i]f any bomber had abstained another pattern of destruction, and hence another event, would have resulted' (852). Under the fine-grained view, every bomber did make a difference, so they can be held responsible for causing the city's destruction.

The problem with Petersson's fine-grained view of causation is its incompleteness. When he defends his view, he acknowledges that the fine-grained view *may* over-stretch the applicability of causal relationships. For instance, the view allows that merely because a pilot took a letter posted from his mother with him, the mother had caused the bombings, which 'would not have occurred in exactly the same manner' (853) if the mother had not posted the letter. To avoid this kind of implications, Petersson contends that 'counterfactual dependence is merely supposed to be necessary for causal involvement, not sufficient' (854), so the fine-grained understanding does not necessarily imply that anything that makes any minute difference to an event causes the event. In this sense, Petersson also rejects P2, which treats counterfactual dependence as a necessary and sufficient condition of causation. Petersson's response is plausible, but it also leaves open the question of what other conditions there are to causation. In other words, because his definition of causation is incomplete, there is no way to test whether the fine-grained view captures our understanding of causation adequately. As such, a theory of complicit responsibility based on the fine-grained view will carry an external theoretical risk until the other conditions are clarified. This risk constitutes a theoretical disadvantage to Petersson's solution to the problem of overdetermination.

To conclude, inclusive ascription blames the innocent, Gardner's criterion of causality assigns causal links too loosely, and Petersson's fine-grained view is incomplete. For these reasons, I argue that Kutz's, Lawson's, Gardner's, and Petersson's solutions to the problem of overdetermination are flawed. The next section will propose a solution that does not contain the problems discussed above.

Contribution as Risk Taking

Based on the previous section, a solution to the problem of overdetermination should meet the following five requirements: First, it should save P4 – it should explain how individuals can be held responsible in cases of fully overdetermined collective harm. Meeting this requirement can fulfil Kutz's metaethical commitment to preventing collective harms by providing individual agents with the right moral reasons. Second, unlike the theory of inclusive ascription, the ideal solution should not blame people for identity expression and wishful contribution. Third, as Gardner and Petersson insist,

the ideal solution should have some causal element, so causality remains central to our moral assessment. Fourth, unlike Gardner's criterion of causation, the ideal solution should be based on a plausible definition of causation. I presume that the counterfactual dependence view of causation is sound. Fifth, unlike Petersson's fine-grained view, the ideal solution should be based on a complete definition of causality.

In the following, I will argue that an individual is complicitly responsible for a collective wrongdoing if she has taken an action that carries an impermissible risk of causing the collective wrongdoing's success, and this action expresses a culpable choice. I will also show that this conception of complicit responsibility can overcome the problem of overdetermination and meet all of the above five requirements.

Guiding actions in an uncertain world

According to Wallace's account of individual culpability, an unjustified action is necessary for holding an individual responsible. Thus, when a person is held responsible for a harm, the primary object of responsibility is still her action. She becomes responsible for the harm only because her action actually causes the harm, and she is responsible for her action – action and harm are not mutually exclusive objects of moral responsibility. Here, the causal relationship between her action and the harm indicates that her action is (pro tanto) unjustified. This analysis, however, also shows that harm is not necessarily the only object of individual responsibility; acknowledging that an action can be unjustified because it causes harm is consistent with accepting that an action can be unjustified for other reasons. This route to saving P4 is attempted by the theory of inclusive ascription, which holds that an action can be unjustified just because it is teleologically explained by the agent's conception of a collective wrongdoing. I have rejected the theory of inclusive ascription, but I will still argue that individual actions within the context of fully overdetermined collective harm can be unjustified for reasons other than actually causing it. I will show that these individual actions can be unjustified because they carry impermissible risks.

The need to attend to risks in the moral assessment of actions arises from the action-guiding nature of morality: Morality is supposed to guide how agents should make their choices *when*, not after, they make their choices. If the assessment of an action is based on conditions that the agent cannot know at the time of the decision, then morality will fail to be action-guiding. In Lepora and Goodin's (2013) words, 'morality must assess the situation in terms of what can be known by the actor at the time she acts,' for 'the privilege of hindsight is not permitted to moralists trying to give guidance as to what people should do at the time of their acting' (61).

Here, risks sometimes matter to whether an action is unjustified because we live in an uncertain world. Although we can know the consequences of a past action with the benefit of hindsight, very often the consequences of an action cannot be known with certainty when an individual makes her choice. Many of our actions are like throwing a paper aeroplane into the air rather than dropping a sandbag from our chest: At the moment of action, the best we can know is very often its potential consequences but not any definite outcome. Against this background, risk taking means taking actions when their outcome is not known with certainty. Although I will not follow Hansson's (2013) terminology in this thesis, it is still helpful to understand different levels of risk taking. According to Hansson, we take 'risks' (12) when we take an action with a known set of specific outcomes, each with a known probability, but we lack the

knowledge of which outcome will have resulted. We encounter 'uncertainty' (12) when we further lack the knowledge of each outcome's probability. We also sometimes act without knowing all the possibilities. In particular, multi-agent interactions, primarily malevolent actions, strongly resist predictions because 'the effects of your actions are combined with those of others, giving rise to compound effects that do not follow from your own actions alone' (16). As such, even if the world is deterministic, we still sometimes face unsurpassable epistemological barriers to knowing the consequences of our actions with certainty. In other words, regardless of whether the world is deterministic, the consequences of some actions are practically indeterminate from the perspective of acting agents – we must sometimes take risks.

If an action is taken amid uncertainty, but its moral assessment still depends on its actual consequences, morality will fail to be action-guiding – it will ask moral agents to make the 'right' choices when it cannot be known which choices are right. Therefore, to preserve the action-guiding nature of morality, the moral assessment of actions taken amid uncertainty should not be based on their actual consequences; instead, I suggest that their moral assessment should be based on their risks, namely their potential consequences in the world. This way, the moral assessment of actions will not be based on (future) conditions that the agent cannot know at the time of the decision.

I will use two cases to show that our moral practices do assess actions based on the risks they carry. The first is Kumar's (2003) case of a drunk driver, who swerves along the street while you are taking a stroll nearby. Although nothing happens to you, your life has been put at risk. Here, Kumar suggests that 'there is nothing suspect about the claim that one has been wronged by the drunk driver' (103). If we assess the driver's drunk driving only based on what has happened, the fact that you survive unharmed cannot explain why drunk driving is unjustified. But if we also consider the risks of drunk driving – it could kill or seriously injure you – this element shows why drunk driving is unjustified. The second case comes from Wallace (1994): 'If I set out to harm someone, but fail because, say, the knife slips out of my hand, I am no less blameworthy (from the moral point of view) than I would have been had I succeeded in inflicting harm' (128). Again, to explain why a missed stab is unjustified even when harm is absent, we must resort to the risk involved in taking the stab – it could have harmed someone.

Here, a critic may doubt if risks really play any role in explaining why some actions are unjustified. It may be argued that because all actions carry risks (in the sense of uncertain consequences), the existence of risks does not distinguish permissible actions from unjustified actions. This criticism is misplaced because I do not mean that the existence of risks per se makes an action unjustified. What I want to argue is that assessing an action requires us to investigate its risk-taking element, and the action is (pro tanto) unjustified if this risk-taking element is impermissible. So, when is an action's risk-taking element impermissible?

To assess the risk-taking element of an action, the assessor should first list all of its major foreseeable potential developments based on information available to the acting agent at the time of action. Then, based on these major developments, the assessor can identify the good and/or bad risks that the action carries. An action carries a good risk when it may cause something intrinsically valuable to happen. For instance,

hanging a lifebuoy near the sea carries a good risk of saving someone's life one day. An action carries a bad risk when it may cause something bad to happen, such as violating someone's rights or destroying something intrinsically valuable. For example, drunk driving carries a bad risk of gravely injuring somebody. Moreover, depending on the level of uncertainty of the action's consequences, the assessor may also estimate the likelihood of different risks. With this information in mind, the assessor should then make a balanced judgment based on the action's good risks, bad risks, and their relative likelihood to determine whether the risk-taking element of the action is permissible. In general, an action carries impermissible risks if the probability-adjusted net value of its risks is negative.

A critic may further argue that such a comprehensive risk assessment is unnecessary for assessing the permissibility of an uncertain action. It may be proposed that an action is unjustified if it *may* cause something bad to happen. For example, a critic may suggest that drunk driving and attempting to stab someone are unjustified just because they may gravely injure someone – an assessor does not need to resort to a comprehensive risk assessment to show that they are unjustified.

However, I argue that this alternative way of assessing the permissibility of an uncertain action will lead to counterintuitive implications. Intuitively, it is permissible for a properly trained driver to drive safely, abiding by all traffic rules. Yet, even the most skilful and safe drivers *may* gravely injure someone despite its unlikelihood. If we insist that a mere possibility of causing something bad to happen is sufficient to make an action unjustified, we must conclude that it is unjustified for a properly trained driver to drive safely – every driver on the road would be committing wrongdoings. This counterintuitive conclusion rejects the alternative way of assessing the permissibility of an uncertain action. On the contrary, a comprehensive risk assessment can explain why safe driving is permissible: Although safe driving carries some improbable bad risks of harming someone, it also carries the good risk of transporting the driver and/or the passenger to their destination, which then enables them to engage in valuable activities. And since the probability-adjusted net value of safe driving is not negative, it is permissible. Therefore, it is not the existence of bad risks per se that explains why drunk driving and attempting to stab someone are unjustified; they are unjustified because the likelihood and seriousness of their bad risks outweigh the value (if any) of their good risks.

Another example that shows the advantage of a risk assessment over the alternative proposal is the permissibility of our different behaviour during the COVID-19 pandemic. Suppose we were back in the UK in January 2021, when the monthly number of cases of COVID-19 confirmed deaths was at its highest. Against this background, it is intuitive that it would be wrong for a person who was experiencing COVID-19 symptoms to randomly walk into an elderly care home without wearing any protective equipment; it is also intuitive that it is permissible for a person who experienced no symptoms and had no recent close contact with people with COVID-19 to visit a supermarket to swiftly buy food for a week. Comprehensive risk assessments could explain both intuitions: The former case involved no good risks but involved bad risks of spreading the disease to vulnerable elderly people and even causing their death, so a balanced judgment would suggest that the visit's risk-taking element is impermissible. The latter case still involved the same bad risks, but their likelihood was much lower than in the former case, and there was a good risk (almost certain indeed)

of enabling the survival of the supermarket visitor, so a balanced judgment would suggest that the visit's risk-taking element is impermissible.

Nonetheless, the alternative proposal would suggest that both cases were impermissible: In the supermarket case, a symptomless person *may* still carry COVID-19, she *may* come close to a vulnerable person in the supermarket, this contact *may* allow the disease to spread, and this disease *may* kill the vulnerable person. Against, if we insist that the mere possibility of causing someone's death is sufficient to make an action unjustified, we must conclude that it would be unjustified for any UK resident to visit the supermarket in January 2021. This counterintuitive conclusion, again, shows why the alternative proposal is inadequate. Therefore, we must resort to comprehensive risk assessments to decide whether an uncertain action is unjustified.

Risking the successes of collective wrongdoings

I have argued that an individual action can be unjustified because it involves an impermissible risk-taking element, and this element should be assessed based on a comprehensive risk assessment. Next, I will show how this theory allows individuals to be held responsible in the context of fully overdetermined collective harms.

Even if a person's action has made no difference to the success of a collective wrongdoing, it could still have made a difference if things happened differently. Expressed in possible worlds semantics, even if the collective outcome would not have been different in the closest possible world (to the actual world) where the person had acted differently, there could be a difference between the resulted collective outcomes in another possible world where the action was taken and the closest one where the action was not taken. In other words, an individual action that has made no difference to a collective outcome may nonetheless involve risks of causing the success of the collective wrongdoing. Then, since the collective outcome is bad, the risk of causing it is also bad, and unless the action involves some good countervailing risks, its risk-taking element will be impermissible. As such, the individual can be held responsible for unjustifiably risking the success of the collective wrongdoing.

Let's examine the aerial bombing of Dresden from the perspective of risk taking. I assume that the pilots' actions fully overdetermined the destruction of Dresden, so none of them had actually caused the destruction. However, this assumption does not imply that none of them had taken the risk of causing Dresden's destruction when they carried out the mission. There were many ways that their dropping of explosives and incendiaries could have made a difference. For instance, there could be a residential area where all but one plane had hit, and then the crew on that plane would have caused the death of many innocent citizens. And many other similar possibilities were foreseeable before the bombings began. All these bad risks matter to the moral assessment of their participation in the mission. A pilot might defend his risk taking by citing the bombings' potential effects on Allied Victory; however, at least, the absence of an actual causal relationship between his action and Dresden's destruction is not an adequate defence. In short, even if the pilots cannot be held responsible for causing Dresden's destruction, they can be held responsible for unjustifiably risking the city's destruction.

In the previous section, I have criticised Kutz's definition of 'contribution', '(intentional) participation', and 'playing a part' as any individual action teleologically explained by a conception of collective action. Now, having shown how individual actions can be unjustified because they risk the success of collective wrongdoings, I propose the following definition of 'contribution': an agent's action contributes to a collective wrongdoing if and only if the action will, on balance, promote the success of the collective wrongdoing, meaning the action has a positive expected utility to the collective goal. An action's potential contribution to a collective wrongdoing constitutes a bad risk that counts towards the impermissibility of the action's risk-taking element. This potential contribution thereby constitutes a (defeasible) reason against taking the action, and thus allows the individual to be held complicitly responsible for the collective wrongdoing.

There are two things to note about this definition. First, not all actions that *may* cause the success of a collective wrongdoing count as contribution. Some actions involve both positive and negative potential effects on collective goals, and not all these actions have a positive expected utility to the goals. Suppose some highly trained rebellions decide to assassinate the commander of an extermination camp. They have a good chance of success, and the genocidal programme will be slowed down considerably if they are successful. However, due to an unlikely accident, the assassination fails, and they end up leaking hiding locations to the commander, thus speeding up the genocidal programme. Even though it is foreseeable that the assassination may facilitate the genocidal programme, the assassination has a negative expected utility to the genocidal goal, so the rebellions' actions do not count as contribution to the genocide. I also believe it is intuitive that they do not.

Second, contributing to collective wrongdoings is not necessarily unjustified. An action contributing to a collective wrongdoing may involve some good risks that make its risk-taking element permissible. For example, transporting ammunition to an extermination camp does contribute to its genocidal goal, but it may be permissible for a person to do so if this is the only way she can save her family from becoming the target of the genocide. It is also not necessary that an action involving impermissible risks of contributing to a collective wrongdoing is unjustified. For instance, the risks of saving the life of a dying soldier may be impermissible if she has committed war crimes and is likely to commit war crimes again if she survives; however, if the doctor treating the soldier has professional obligations or has sworn an oath to save any lives regardless of their past and future, these reasons may outweigh the risks of saving the soldier and make saving her a permissible (or even obligatory) act. As such, an action that contributes to a collective wrongdoing is only pro tanto unjustified; whether it is overall unjustified depends on its other potential consequences and other moral considerations.

Solving the problem of overdetermination

At the beginning of this section, I have listed five requirements for an ideal solution to the problem of overdetermination. Satisfying these five requirements gives a solution significant advantages over Kutz's, Lawson's, Gardner's, and Petersson's solutions. In the following, I will explain how holding individuals complicitly responsible for risking the successes of collective wrongdoings meets these five requirements.

First and foremost, my proposal saves P4. It allows individuals to be held responsible for taking impermissible risks even when the outcome of the collective wrongdoing is actually fully overdetermined. This way, this proposal satisfies Kutz's (2000) commitment that 'a theory of individual accountability must provide individual agents with reasons to avoid and reasons to respond to even marginal participation in harmful acts' (123), thus 'controlling the essentially collective harms that threaten individual interests' (123).

Second, my definition of contribution does not count identity expression and 'wishful contribution' as contribution to collective wrongdoings. As identity expression does not necessarily involve the risk of causing the successes of collective wrongdoings, it does not necessarily qualify as contribution to collective wrongdoings. In the example that a person privately changes her diet to show her solidarity with a radical animal rights organisation, this change involves no risk of causing the success of the organisation's violent missions, so my definition of contribution does not count her dietary changes as contribution to the violent missions. This way, this definition avoids the counterintuitive implications of the theory of inclusive ascription in cases of identity expression. On the other hand, wishful contribution has, by definition, no chance of causing the success of collective actions; it is just falsely believed by the agent to be so. Therefore, my definition does not count wishful contribution, such as praying for the success of a genocidal programme, as (real) contribution, and thus avoids inclusive ascription's counterintuitive implication that individuals can be blamed (and punished) for wishfully contributing to collective wrongdoings.

Third, my proposal does not abandon causality in assessing complicit responsibility. Although an action does not need to have actually caused the success of a collective wrongdoing to qualify as contribution to it, it must involve the risk of causing its success to qualify as such. In other words, there must be a potential causal connection between the individual action and the collective outcome – causality remains central to our assessment. Therefore, my proposal is consistent with Gardner's and Petersson's insistence on the central place of causality in assessing moral responsibility.

Fourth, my proposal still uses the counterfactual dependence view of causation, so it does not overstretch the applicability of causation like Gardner's criterion of causality. In the example that two teams are ordered to assassinate a drug dealer, the counterfactual dependence view of causation would not imply that Team A, the time bomb of which explodes after the dealer's death, has caused the dealer's death. It is because the dealer would still have died if Team A had not plated the time bomb. On the contrary, Gardner's definition of causation entails that Team A has caused the dealer's death.

Fifth, compared to Petersson's fine-grained understanding of causation, my proposal does not deny the sufficiency of the counterfactual dependence condition to causation. A complete definition of causation gives my solution to the problem of overdetermination an advantage over Petersson's.

Conclusion

The chapter began by asking what the conditions of holding civil servants complicitly responsible for past state atrocities are. To investigate this thesis sub-question, I

turned to examine the criteria of complicit responsibility. Wallace argues that a person can be held responsible for her action if it is unjustified and expresses a culpable choice. Throughout this chapter, I have argued that an individual action can be unjustified because its risk-taking element is impermissible; this permissibility depends on a balanced risk assessment, and causing the successes of collective wrongdoings is a kind of bad risks. I have also defined contribution to collective wrongdoings as risking their successes. Therefore, I argue that an individual is complicitly responsible for a collective wrongdoing if she has unjustifiably contributed to the collective wrongdoing, and this contribution expresses a culpable choice. As such, I conclude that a civil servant is complicitly responsible for past state atrocities if she has unjustifiably contributed to the atrocities, and this contribution expresses a culpable choice.

The above conclusion also rejects Murphy's (2017) last challenge to retributivism's applicability in transitional circumstances. Murphy argues that retributive theories cannot individuate individual responsibility for collective wrongdoings based on the assumption that individual responsibility must be assessed based on 'what an individual caused on his or her own, consequences that are assumed to be under his or her control' (90). However, this assumption is invalid because individual responsibility can also be assessed based on the risk-taking element of the individual action. I therefore conclude that Murphy's all four challenges to retributivism's applicability in transitional circumstances have failed.

Chapter 4
Justifications and Excuses

This is the final chapter that investigates the thesis question of whether civil servants should be punished for past state atrocities in transitional circumstances. In the previous chapters, I have defended the view that retributivism is applicable in transitional circumstances, and that it offers a defeasible justification for punishing responsible perpetrators of atrocities proportionally. Thus, the answer to the thesis question depends partly, but still crucially, on whether civil servants can be morally responsible for state atrocities. I have also developed an account of complicit responsibility in the context of collective wrongdoings: If an individual has contributed to a collective wrongdoing, meaning her action has a positive expected utility to the collective goal at the time of action, then this contribution is pro tanto wrong even if it does not cause the collective outcome in the end. She then bears complicit responsibility for the collective wrongdoing if her contribution is overall unjustified, and the contribution expresses a culpable choice.

These discussions present a preliminary answer to the thesis question: A civil servant should be punished for atrocities if she bears complicit responsibility for them and if other forward-looking considerations during transitions do not defeat the retributivist justification. Some may expect a simple ‘yes’ or ‘no’ to the question, but the commitment against guilt by association requires each civil servant’s responsibility to be determined individually. The judgment of their culpability will be unfair if it ignores their individual actions, mental states, and contexts. A philosophical discussion can, at best, spell out the conditions of justified punishment and let practical institutions determine ‘yes’ or ‘no’ for each civil servant.

The conditions of complicit responsibility may suggest that most civil servants of states engaging in atrocities should be punished – supposing the government is largely rational, most civil servants it hires will likely provide positive utility to its goals. However, the conditions of justified punishment may be a lot more restrictive than what this picture suggests, for civil servants can claim a myriad of justifications and excuses to defend their innocence.

Justifications and excuses counter blame against individuals in different ways. When an agent defends her accused action with a justification, she does not deny that she has chosen to do that; instead, she justifies her action by arguing that it is morally permissible or, at least, not as severe as the accuser claims. A common justification that civil servants employ is the lesser evil justification. It is employed when an accused civil servant claims that the state atrocities would have been more successful if she had not performed her accused role. The truth of this counterfactual claim will imply that what she has done in her role does not qualify as contribution to the atrocities, so what she has done is not unjustified. One example of the lesser evil justification was why Hans Globke, a Chancellor of West Germany, was not perceived as culpable for being ‘an official commentator on and interpreter of the 1935 Nuremberg racial laws’ (Elster, 2004, 146). This was because he was widely believed to have given mild interpretations to the racial laws to prevent the worst from happening (although, according to Elster, some historians also pointed out that Globke

had written a directive to prevent Jews from changing their names to disguise their Jewish descent).

On the contrary, when an agent defends her accused action with an excuse, she does not deny that the action is unjustified; instead, she argues that she is not responsible for it because it does not express a culpable choice. In this sense, civil servants can claim to be 'half-souled' – lacking full agency for what they have done, and thus not morally responsible for their action. A common excuse that civil servants use is the excuse of duress. It is used when an accused civil servant claims that she has performed her role only because refusing superior orders would have led to (serious) consequences. For example, Mulka, an Auschwitz guard brought to trial in the 1960s, defended himself claiming that he would have been killed had he disobeyed (Elster, 2004, 153). The court acknowledged that duress could be an excuse, but the evidence presented to the court suggested that disobedient camp officers had not been killed but only transferred. Mulka was sentenced to 14 years in prison in the end.

To better understand the conditions under which a civil servant can be justifiably punished, this chapter will investigate the final thesis sub-question: What justifications and excuses can civil servants claim to avoid blame and punishment? In the following, I will grant the validity of the lesser evil justification and the excuse of duress and turn to examine four more morally puzzling justifications and excuses: the cog-theory justification, the superior order justification, the excuse of absent discriminatory intent, and the excuse of factual and moral ignorance. Among them, I will argue that only the excuse of ignorance is valid, albeit with qualifying conditions.

The Cog-Theory Justification

I start with the cog-theory justification. It takes the following form at its core: If I had not done it, someone else would have, and the collective outcome would not have been different. The cog-theory justification is sometimes called the 'fungibility excuse' (Elster, 2004, 144) or 'excuse from alternative cause' (Thompson, 1980, 909; Boot, 2019, 140). But this defence is not an excuse because the object of contestation is not the civil servant's state of mind; it has nothing to do with the quality of the choice expressed by the action. Instead, the defence is a justification as it contests the wrongness of the action. It attempts to show that the action is not unjustified because no difference to the collective outcome has been made.

The validity of the cog-theory justification is significant because it was commonly claimed in post-war trials. Eichmann defended himself by claiming he was 'after all only a "tiny cog" in the machinery of the Final Solution' (Arendt, 1994, 289). The defence was also, perhaps with slight exaggeration, 'what all the defendants in the post-war trials said to excuse themselves' (Arendt, 2003, 29). The defence has also been claimed in international courts. For instance, Thomas Lubanga, a member of the Union of Congolese Patriots who was convicted of recruiting and using child soldiers by the International Criminal Court (ICC) in 2012, defended himself by arguing that he had made no difference, as his role was 'in no way determinative of the crime of recruiting child soldiers' (Stewart, 2012, 1192). As Stewart notes, the ICC implicitly acknowledges the validity of the cog-theory justification. He writes that existing international criminal law requires the guilty individual to have made an essential

contribution to the crime; the accused would not be convicted unless her action was causally necessary to the criminal outcome. Lubanga was still convicted by the court, albeit on the grounds that his contribution to the crimes was essential.

On what grounds does the cog-theory justification counter the blame against civil servants? To begin, it assumes the background of a political system that functions like a machine: 'Each cog, that is, each person, must be expendable without changing the system' (Arendt, 2003, 29). Thompson (1980), similarly, describes that 'in organizations persons often are fungible' (909). In other words, the cog-theory describes the general phenomenon that when a civil servant stops performing her role, the bureaucracy is structured in a way that allows her role to be taken up by another person immediately, and then the state machine can keep running without any issue. This phenomenon justifies the counterfactual claim that even if a civil servant had not performed her role, the atrocity outcome would not have differed. Then, on the assumption that a person can be held morally responsible for a collective wrongdoing only if she has caused the collective outcome (making a difference to it), no expendable civil servant can be held responsible for state atrocities. If the cog-theory justification is valid, most civil servants will be blameless for executing orders that contribute to atrocities, as '[v]ery few atrocities are so dependent on the acts of any one individual that we can say with confidence that they would certainly not have transpired absent any one accused's individual agency' (Stewart, 2012, 1190).

I reject the validity of the cog-theory justification. I reject it for the same reasons as, in the previous chapter, I reject the conclusion that no participant of an overdetermined collective wrongdoing can be held morally responsible. To briefly recapitulate, an individual is complicitly responsible for a collective wrongdoing if she has unjustifiably contributed to the collective wrongdoing, and this contribution expresses a culpable choice. Even if a civil servant's action has not made any difference to the collective outcome, it can still qualify as contribution to the collective wrongdoing if it has a positive expected utility towards the goal. Many possible developments may follow a civil servant's refusal to execute state orders. Given the state's bureaucratic structure, the possible world closest to the actual one will see another person perform her role equally well as she does in the actual world. This grounds the truth of the counterfactual claim. However, there are still possible worlds in which her refusal causes the state atrocities to be less successful: Her role may not be taken up for a significant period due to labour shortage, her role may be taken up by another civil officer who performs the role a lot poorer, and her refusal may inspire many others to do the same. Compared to these possible developments that negatively affect the state's atrocity goals, her obedience does make a difference. As such, when we calculate the expected utility towards state atrocities of her execution of orders, it would score a zero had the cog-theory counterfactual been true, but it would deliver positive utility to state atrocities had the above worse-case scenarios been true. In other words, although her execution of orders has made no difference to the atrocity goals, it has still risked causing their successes. Then, unless other countervailing considerations outweigh the pro tanto wrongness of her contribution, she can be held complicitly responsible for the atrocities despite the truth of the cog-theory counterfactual. Hence, I conclude the cog-theory justification is inadequate to show that a civil servant's action is not unjustified.

The Superior Order Justification

Unlike the cog-theory justification, the superior order justification does not contest whether the civil servant's contribution is pro tanto wrong. Instead, it proposes a moral reason that justifies the contribution. It suggests that civil servants have a professional obligation to obey superior orders, including those conducive to atrocities. I will argue that the superior order justification is invalid in the following.

Civil servants, most prominently soldiers, have frequently claimed the superior order justification. In the post-war trials of Nazi officers, the defendants repeatedly argued that they were obliged to give 'obedience to superiors as well as obedience to the laws of the land' (Arendt, 2003, 56). Before the International Criminal Tribunal for the former Yugoslavia (ICTY), both Serb and Croat defendants argued that 'I was only following orders' (Insco, 2003, 389). A slightly different case is the due obedience law in Argentina, passed in June 1987 under the threat of a military coup. It created an irrebuttable presumption that all officers subordinate to the top leaders of the military junta were 'innocent because they were following orders' (Crawford, 1990, 27). International courts have also recognised the superior order justification to a limited extent. As McCoubrey (2001) observes, both international military tribunals at Nuremberg and Tokyo allowed the superior order defence to mitigate (but not exculpate) punishment in some circumstances. This treatment of the defence was then followed by the ICTY and the International Criminal Tribunal for Rwanda (ICTR).

The idea that civil servants have a professional obligation to obey superior orders can find support from the conventional view that civil servants are governed by 'the ethic of neutrality' (Thompson, 1985, 555), meaning that they should obey the orders of the state and become 'a completely reliable instrument of the goals of the organization, never injecting personal values into the process of furthering these goals' (556). This view can, at least, be traced back to Max Weber's (2014) model of bureaucracy, according to which: 'The honor of the civil servant is vested in his ability to execute conscientiously the order of the superior authorities, exactly as if the order agreed with his own conviction. This holds even if the order appears wrong to him and if, despite the civil servant's remonstrances, the authority insists on the order' (95). For Weber, civil servants should obey their superiors even if judgments diverge between them.

Why do civil servants have a professional obligation to obey superior orders? It is because they have promised to do so. As Thompson (1985) explains the ethic of neutrality, civil servants are 'morally obliged to serve the organization in this way because their acceptance of office is voluntary: it signifies consent' (556). In other words, civil servants' professional obligation is in essence a contractual obligation: By voluntarily signing an employment contract and/or swearing an oath to serve the state, they acquire an obligation to obey superior orders to serve the state. It is like when we promise a friend to bring her lunch, we owe her an obligation to bring her lunch; we would wrong her if we failed to do so.

I think the contractual relationship between civil servants and the state explains why they have a professional obligation to execute orders to promote normal state functions, such as protecting the people and arranging the provision of public goods. However, in the following, I will suggest two reasons why this contractual relationship does not justify any professional obligation to obey superior orders conducive to atrocities.

Illegal superior orders

The first reason why civil servants cannot use the superior order justification to defend their contribution to atrocities is conditional: If a civil servant's contract with the state requires her to obey the law, and the atrocity-inducing order she receives is illegal, then her contractual relationship with the state does not generate any obligation to obey the illegal superior order.

It is common for civil servants to have consented to obey the law and/or promised to exercise their power within the scope of the law. For example, in August 1934, Germany enacted a law on the attestation of officials and soldiers. It required public officials to swear this oath of service: 'I swear: I shall be loyal and obedient to Adolf Hitler, the Fuehrer of the German Reich and people; respect the laws, and fulfill my official duties conscientiously, so help me God' (Office of the United States Chief Counsel for Prosecution of Axis Criminality, 1946). And even if the contractual content does not explicitly require the civil servant to obey the law, but only requires her to obey the state, this requirement to obey the state still implies that the civil servant is obliged to obey the law. It is because the will of modern states is represented by their laws and, more fundamentally, their constitutions. Obeying illegal superior orders will therefore constitute a failure, instead of a success, to comply with the demands of the state.

I have also argued in Chapter 2 that not all instances of committing atrocities are legal at the time of their commission. In states that have ratified the Rome Statute of the ICC, it would be illegal to commit, order, facilitate, or contribute to atrocities, so superior orders that command civil servants to assist with atrocity goals would be illegal. More generally, regardless of a state's status of ratification of international treaties, as long as it has criminalised elements of atrocities – such as murder, assault, enslavement, torture, and sexual violence – superior orders that command civil servants to commit the state's criminalised elements in the process of carrying out atrocity goals would also be illegal.

I do not mean that civil servants who have contributed to state atrocities must have promised to obey the law, nor do I mean that the atrocity-inducing orders they receive must be illegal. The above reason against the existence of civil servants' professional obligation to obey atrocity-inducing orders is, after all, conditional. Yet, when the right conditions obtain, the superior order justification is invalid, and it remains so even if the second reason that I will introduce next is flawed.

Bad promises

The second reason why civil servants do not have any professional obligation to execute atrocity-inducing orders is that bad promises do not obligate. Promises do not always create obligations, for the moral quality of the promised act can prevent the promise from obligating the promisor. Simmons's (1976) example for this point is this: 'I make two promises to a friend – one to help him commit murder most foul, the other to give him half my yearly income' (284), and 'while both promises are real promises, the latter obligates me while the former does not' (284). In other words, the act of promising alone is not sufficient to obligate the promisor; the act of promising must be coupled with the permissible nature of the promised act to be jointly sufficient to obligate the promisor.

In normal circumstances, the content of civil servants' promises to the state is usually permissible. Their work may even promote citizens' fundamental moral interests when they fulfil orders to serve state functions such as securing peace, resolving property rights disputes, maintaining the rule of law, administering distributive justice, and providing public goods. In these cases, their promises to the state do ground their professional obligation to obey superior orders. However, once their superior orders command them to take impermissible actions that contribute to atrocities, their promises become bad promises, and they cease to be bound by any contractual obligation to fulfil the orders. As a result, civil servants cannot use their normally applicable professional obligation to justify their execution of orders that contribute to atrocities – the superior order justification is invalid.

This second reason for the invalidity of the superior order justification is not conditional upon whether the civil servant has promised to obey the law or whether her contribution has violated any national or international law. Even if she has sworn allegiance to only the leader but not the law or the state, her promise will lose obligatory force once the content of the promise becomes impermissible. Therefore, even if the illegality of the superior order does not block the superior order justification, it will still be blocked by bad promises' failure to obligate their promisors.

Here, a critic may point out that it is counterintuitive that every civil servant who has carried out some impermissible orders deserves to be blamed and punished. But this is not a necessary implication of rejecting the superior order justification. While rejecting this defence implies that civil servants cannot justify their contribution to atrocities by appealing to any obligation to obey superior orders, they can still appeal to other justifications and excuses to show that they are not blameworthy. For example, they can excuse themselves by claiming that they do not know the illegality and impermissibility of their superior orders, thus appealing to the excuse of ignorance (which I will acknowledge as a valid excuse in a later section). Therefore, rejecting the superior order justification does not imply that civil servants who have carried out impermissible orders must be wrong or deserve punishment for doing so.

The Excuse of Absent Discriminatory Intent

Claims of excuses acknowledge the overall wrongness of the accused action but contend that it does not express any culpable choice. This section will examine the excuse of absent discriminatory intent and argue that it is invalid. At its core, this excuse assumes that a civil servant's choice to contribute to atrocities is culpable only if she shares discriminatory intent with the state. Discriminatory intent contains the specific purpose of inflicting atrocities on a targeted population. Personal purposes such as earning a salary and advancing one's career are not discriminatory. When a civil servant appeals to the excuse of absent discriminatory intent, she declares that her motive for contribution is entirely personal; she has never wished the atrocities to occur. 'I have no evil intention', she says.

The significance of the excuse of absent discriminatory intent can be shown by the fact that civil servants' intentions behind staying in their job are quite diverse, and many of them execute their orders for purely personal purposes. Elster (2004) observes that not all wrongdoers of atrocities are 'fanatics' (142) motivated by the

state's goals. He also identifies 'opportunists, losers, and the malicious' (141) who are motivated by personal gains, 'conformists' (141) who are afraid of losing a career, and 'the thoughtless' (140) who simply lack judgment and prudence. This observation is shared by Arendt (1994) in the phenomenon she calls 'the banality of evil' (287). Regarding Eichmann, she writes, 'Except for an extraordinary diligence in looking out for his personal advancement, he had no motives at all' (287). This thesis of the banality of evil warns that it was the normal people who lacked evil intentions that became instruments of the worst evils in human history. May (2005) also observes that minor players of atrocities rarely have 'any particular reason to do this act as opposed to hundreds of others open to the agent' (130); and indeed, 'most soldiers and other minor players will willingly do whatever they are commanded to do' (130). In other words, rather than executing *that* order to advance the state's discriminatory goals, they are just executing orders indiscriminately.

Whether these civil servants can be excused depends on whether a culpable choice to contribute to atrocities must be motivated by discriminatory intent. May (2005) argues that proving discriminatory intent is necessary to hold a minor player legally responsible for crimes against humanity. To be fair, I must note that May explicitly states that his focus is on the question of legal instead of moral guilt. Still, I think his case can be reformulated as a moral argument that deserves our attention.

May's (2005) argument begins with the following premise: To punish a minor player for crimes against humanity, she must have participated with a state of mind sufficiently connected to the collective wrongdoing. Here, a sufficiently strong connection is required because 'it cannot be assumed that this individual has a "guilty mind" concerning the larger crime' (127). Then, May identifies two candidates for this connection. The first is the knowledge that one's action contributes to a wrongful group plan to commit crimes against humanity. The second is discriminatory intent – the minor player aims to contribute to the wrongful group plan; she intends to 'do something specific that is proscribed by international law' (130). Next, May rejects that mere knowledge constitutes a sufficiently strong link, 'for that knowledge may be completely incidental to why the person so acted' (126). He therefore concludes that only discriminatory intent can sufficiently connect minor players to state atrocities.

I agree with the first premise of May's argument: To show that a choice is culpable, it must be connected to the wrongdoing in a particular way; otherwise, there is no need to investigate the state of mind of the person we attempt to hold responsible. However, I disagree with May regarding what constitutes a sufficient connection. May thinks mere knowledge is not strong enough to be the link because the link cannot be incidental to why the minor player contributes to the state's discriminatory goals. In other words, he assumes that a culpable choice to contribute involves a 'positive' requirement: The person's choice must not only be made with the knowledge that it wrongly contributes to atrocities, but it must also be motivated by that precise knowledge. On the contrary, I argue that a culpable choice to contribute only involves a 'negative' requirement: The person knows that her action will wrongly contribute to atrocities, but this knowledge fails to demotivate her to avoid that choice. As Wallace (1994) explains, underlying our practice of holding people responsible for violations of moral obligations is 'a commitment to the existence of justifications that support those obligations, specifying reasons that can motivate compliance with them' (119). In other words, given that a person is aware of the justifications against an action, these

justifications *should* motivate her to choose to comply with her obligation. Thus, contrary to May, I argue that knowledge constitutes a sufficient connection. A civil servant's choice to contribute can be culpable even if she lacks discriminatory intent – even if her choice to contribute is not motivated by the purpose of inflicting atrocities. Her choice is culpable as long as she knows that her action will wrongly contribute to state atrocities. Hence, I conclude that the excuse of absent discriminatory intent is invalid.

Furthermore, holding on to the positive requirement of culpable choices will lead to counterintuitive implications. There is a paradigm type of villain that we commonly recognise: The selfish, cold politician, capitalist, or crime boss, who places her self-interests above the rights and welfare of everyone else and pursues what she wants by all means even if they require 'deep sacrifices' from others. 'I don't care,' she says. The motives of this type of villain are entirely person; their actions are not motivated by their knowledge of how their actions will wrong other people. Still, it is intuitive that they are blameworthy for taking those actions. If we hold on to the positive requirement of culpable choices and assume that they must be motivated by knowledge of wrongdoing, then these villains will not be culpable for their wrongdoings. Instead, if we accept the negative requirement and assume that culpable choices only require a failure to be demotivated by knowledge of wrongdoing, then these villains can be held responsible for their wrongdoings.

I will respond to three objections before I turn to the next section. First, some may point out that we do sometimes grant excuses based on the personal intentions behind people's commission of wrongdoings, and it seems that my account fails to recognise this possibility. It does not. By denying the excuse of absent discriminatory intent, I mean that claiming that one's intention is entirely personal cannot *automatically* excuse the accused – it is not necessary that the choice of civil servants who contribute with purely personal purposes is non-culpable. My account is compatible with recognising that some claims of personal intentions do counter blame when they appeal to valid excuses or justifications. A civil servant may claim she thought she was doing good because she did not know about the atrocities. This appeals to the excuse of ignorance. She may say she just intended to avoid punishment for disobedience. This appeals to the excuse of duress. She may appeal to the lesser-evil justification by saying that she intended to prevent the worse from happening. When these claims of personal intentions counter blame, they counter blame because they appeal to valid justifications and excuses, not because they are not discriminatory against the targeted population.

Second, May might respond that I have missed his argument: What justifies the necessity of discriminatory intent is not only to show that the civil servant's choice is culpable, but also that she is culpable enough to deserve punishment. In other words, May might be appealing to the proportionality requirement of retributivism to justify the high bar of punishment he sets. As the argument goes, because contribution made with mere knowledge is not blameworthy enough to be proportional to deprivations of liberty, mere knowledge does not constitute a sufficient link to justify punishment; only discriminatory intent does. Nonetheless, I insist that proving discriminatory intent is not necessary to punish a civil servant. I agree that all other things equal, a person who commits a wrongdoing with evil motives is more culpable than another person who commits the same wrongdoing with mere knowledge of its wrongness. However,

judgments of culpability depend on many other factors, and it is counterintuitive to give 'lexical priority' to discriminatory intent. Suppose there is a messenger who passes genocidal orders with discriminatory motive once a month and an executioner who executes the orders without discriminatory intent every day. It is counterintuitive to insist that the messenger must be more culpable than the executioner, and that punishing the messenger is justifiable while punishing the executor is not. Since civil servants without discriminatory intent can be more culpable than those with discriminatory intent, I reject that the proportionality requirement can justify treating discriminatory intent as a necessary condition of punishment.

Lastly, another way May might respond that I have missed his point is that the 'sufficient connection' should not only link the individual's state of mind to her wrongful contribution but also link the individual action to the state's collective wrongdoing. He might suggest that it is really the latter requirement that justifies the high bar of punishment he sets. And if this latter requirement is not satisfied, individuals cannot be blamed (and punished) *for* state atrocities, and they can only be blamed *for* individual acts such as rape and murder. As I have argued in the previous chapter, it is right there must be some connection between the individual action and the collective wrongdoing to hold an individual complicitly responsible for it; yet this connection is not filled by a mental element. Instead, the connection is filled by the potential causal relationship between the individual action and the collective outcome. It is the risk of making state atrocities more likely to succeed that qualify some civil servants' actions as wrongful contribution. Indeed, if we ignore the potential causal connection and treat discriminatory intent as the sufficient condition for blaming individual actions for collective wrongdoings, we will reencounter the problem of wishful contribution; we will be forced to conclude that a person who secretly prays for the success of genocide is blameworthy for the genocide even if it is impossible for her praying to facilitate the genocide. Therefore, this alternative interpretation of May's argument still fails to defend the necessity of discriminatory intent to justify punishment.

The Excuse of Ignorance

I have argued that a choice is culpable only if it is made with the knowledge that the action is unjustified. This precondition is precisely what the excuse of ignorance challenges. As the validity of the excuse of ignorance is not controversial, I will focus on investigating the conditions of its validity in this section.

Factual and moral ignorance among civil servants

The issue of ignorance is significant because it can be quite prevalent among civil servants. Knowing the moral reasons against an action involves two types of knowledge. The first is moral knowledge, or knowledge of moral principles. The second is factual knowledge. For example, knowing that drunk driving is wrong involves the moral knowledge that gravely risking others' lives is wrong and the factual knowledge that drunk driving gravely risks others' lives. Lacking either knowledge allows the driver to validly claim that she does not know that drunk driving is wrong. Correspondingly, ignorance can be moral and factual, and civil servants can readily provide evidence for either or both types of ignorance.

Regarding factual ignorance, Boot (2019) points out that the 'highly compartmentalized structure of modern bureaucracies' (140) may prevent civil

servants from knowing that their actions contribute to government wrongdoings. A strong division of labour allows civil servants to fulfil highly specialised roles despite being ignorant of broader causal connections. Furthermore, Skerker (2020) points out that organisations' goals are often diverse and general. Members may understand their actions as contributing to some broadly construed goals but 'the size, complexity, and compartmentalization of organizations blind most participants to the specific collective actions the organization performs' (278), especially the wrongful ones.

On the other hand, moral ignorance can be evidenced by the social norms that support the atrocities and the widespread conformity among civil servants. As Arendt (1994) recounts, Eichmann said that 'the most potent factor in the soothing of his own conscience was the simple fact that he could see no one, no one at all, who actually was against the Final Solution' (116). She concludes that he represents a new type of criminal who 'commits his crimes under circumstances that make it well-nigh impossible for him to know or to feel that he is doing wrong' (276). Murphy (2017, 55-60) also observes that wrongdoings in pre-transitional societies are normalised in a sense that they become unexceptional – wrongdoings are not deviations from the norm but the very norm itself. In such circumstances, we cannot straightforwardly assume that civil servants must know the wrongness of their actions.

The conditions of culpable ignorance

If ignorance always excuses wrongdoings, any civil servant who has been factually or morally ignorant can be excused for contributing to atrocities, and the phenomenon described above will imply that a large majority of civil servants should not be punished. However, this is not the case, for ignorance does not always excuse wrongdoings. Ignorance can be culpable, and culpable ignorance does not excuse – this is why we sometimes counter someone's excuse of ignorance by saying, 'But you should have known that'.

Under what conditions is ignorance culpable? The first thing to note is that no one can be *directly* culpable for ignorance. As Zimmerman (1997) and Rosen (2004) argue, we cannot directly control our beliefs. We can take actions – observe, read, listen, and contemplate – to influence our beliefs, but changes in what we believe to be true just *happen*. In other words, unlike actions, our beliefs are not subject to our choices. They are not subject to direct control like our voices and limbs. And since we can only be directly responsible for things subject to choices, we cannot be directly responsible for our ignorance.

Still, one can be *derivatively* culpable for ignorance. As Rosen (2004) proposes, ignorance can derive its culpability from some prior culpable acts or omissions that cause the ignorance. These acts or omissions are failures of one's 'procedural epistemic obligations' (301), which refer to obligations 'to ask certain questions, to take careful notes, to stop and think, to focus one's attention in a certain direction, etc.' (301). Suppose a driver hit someone on a zebra crossing because she looked at her phone before she reached the crossing. She did not know that someone was on the crossing, but this ignorance would not excuse her failure to press the brake, for this ignorance was culpable due to her violation of the epistemic obligation to keep an eye on the road, especially before she reached the zebra crossing.

The possibility of culpable ignorance implies that civil servants cannot excuse themselves by claiming factual or moral ignorance per se. A valid excuse of ignorance further requires that the ignorance is non-culpable, and this requires that the ignorance is not caused by prior culpable violations of procedural epistemic obligations. Who can claim a valid excuse of ignorance will ultimately depend on individual circumstances. However, the requirement of non-culpable ignorance shows that a lot fewer civil servants can be excused than what the initial picture of state compartmentalisation and widespread conformity suggests.

Nonetheless, Zimmerman's (1997) and Rosen's (2004) complete account of culpable ignorance will imply that almost all civil servants can claim a valid excuse of ignorance. They have not paid any specific attention to civil servants and atrocities in their articles, but they both argue that culpable ignorance is extremely rare in general. Their argument begins with the assumption that a person can be only culpable for either an akratic act or an act done out of culpable ignorance. This applies not only to the primary wrongdoing in question but also to the failure of epistemic obligations from which the culpability of ignorance is derived. Therefore, the culpable failure of an epistemic obligation must also either be an akratic act or an act done out of culpable ignorance. If it is the latter, '[t]he hunt is still on' (Zimmerman, 1997, 416) until the ignorance is ultimately traced to an akratic act. Hence, all culpable actions must either be an akratic act itself or can be traced to an akratic act; an action is not culpable if it is neither akratic nor can be traced to an akratic act. Here, an akratic act is defined as an act that 'involves a belief on the agent's part that he or she is doing something morally wrong' (421), and this belief must be occurrent rather than dispositional, meaning that the agent must consciously believe that her action is wrong at the time of action (see also Rosen, 2004, 307-9). Both Zimmerman and Rosen then point out that it is infrequent for people to commit wrongdoings, including failures of epistemic obligations, while being conscious of their actions' wrongness. For Zimmerman and Rosen, the rarity of akratic acts shows that wrongdoers can often claim a valid excuse of ignorance – 'culpable ignorance occurs less frequently, perhaps far less frequently, than is commonly supposed' (Zimmerman, 1997, 411).

Contrary to Zimmerman and Rosen, I argue that civil servants can claim a valid excuse of ignorance a lot less frequently than they suggest. The first reason is that identifying akratic acts among civil servants is not as difficult as Zimmerman and Rosen suggest. FitzPatrick (2008) points out that people's experience of guilt and shame can 'constitute perfectly good evidence that they acted akratically' (597), and phenomenon associated with guilt and shame are quite readily observable among civil servants in the context of state atrocities. As Alvarez (1997) notes, 'Many of the agents of destruction … and many of the camp guards and functionaries, suffered mental breakdowns, exhibited physical symptoms of stress such as ulcers, were issued alcohol to numb them to the horror and allow them to sleep, and, some of them, committed suicide' (157). Moreover, when Fanon (1961) recounts his observations as a clinical psychiatrist during the French occupation of Algeria, he writes about European patients who had anxieties due to how they mistreated Algerians. One of them was a European police officer who was found 'leaning against a tree, covered in sweat and having a panic attack' (196) after encountering an Algerian he had tortured in the hospital. This officer visited Fanon because he kept hearing screams at night after doing the interrogations. The stronger these signs of guilt and shame are in the context of atrocities, the more common the akratic acts are among civil servants.

The second reason is that Zimmerman and Rosen have wrongly assumed that a culpable action must either be akratic or done out of culpable ignorance. Zimmerman's (1997) justification for making occurrent belief a necessary condition of a non-ignorant culpable act is that 'one incurs culpability for one's action only if one's belief concerning wrongdoing plays a role in the reason for which one performs the action' (421-422). And since a non-occurrent belief concerning a wrongdoing plays no role in motivating it, a wrongdoing done without an occurrent belief of its wrongness is not culpable. This argument shows that Zimmerman, like May, assumes that a choice must meet a positive requirement to be culpable; he assumes that a culpable choice must be motivated by a belief about the wrongness of the action. However, as I have argued, culpable choices only involve a negative requirement: Despite knowing the action's wrongness, the individual fails to be demotivated by that knowledge and refrain from the wrongdoing. This failure can happen even if that knowledge does not become occurrent when she commits the wrongdoing. Thus, an individual can commit a culpable wrongdoing with neither ignorance nor any occurrent belief about its wrongness. Not all non-ignorant culpable actions are akratic; some can be done with merely dispositional knowledge – knowledge that can be retrieved when needed like students' answers to an exam. Indeed, treating occurrent belief as a necessary condition cannot explain why we sometimes blame people for negligence. Peels (2011, 580-581) gives the example of a babysitter who has been told that the child is allergic to milk. Suppose the babysitter then forgets about the allergy and gives milk to the child. Although the babysitter does not act from akrasia – she is not conscious of the imminent harm she will cause — she is culpable for feeding milk to the child. Only by weakening the culpable choice requirement to dispositional beliefs can our intuitions about negligent cases be accommodated.

For civil servants, dropping the requirement for akrasia implies that they can be culpable for contributing to atrocities and failing epistemic obligations even if they have not been conscious of the wrongness of their wrongdoing. Their choice is culpable if, when cued, they can retrieve why their action is wrong. Therefore, I reject the conclusion that almost all civil servants can claim a valid excuse of ignorance.

Epistemic obligations

I have shown that the culpability of ignorance is derived from prior culpable failures of epistemic obligations, so what epistemic obligations do civil servants have? What level of epistemic effort to investigate their contribution's factual and moral connections shall we expect from them? I am aware that non-culpable ignorance of one's epistemic obligations can excuse her from being held culpable for failing them. But here, I will put forward two principles of epistemic obligations that, I believe, every socialised person across different cultures knows.

The first principle is this: If we know that our action may affect others, we have a defeasible obligation to spend some effort to understand its potential consequences and the moral reasons for and against the action. In our daily life, we make sure that no one is passing through the door when we close it; we anticipate others' reactions to our words before we say them; we think twice when a family member asks us to do something at a time that we have promised to meet our friend; and we expect others to do the same under similar circumstances. These practices are all guided by the

principle I outline above, and they serve to show that the first principle is universally recognised.

The second principle is this: The more consequential we know our actions can be on others, the stronger the obligation is on us to acquire factual and moral knowledge about the action. Adjusting the above examples, if the door is a thick iron door rather than a regular wooden door, we will be more careful when we close it; if we are speaking to a victim with some traumatic experience, we will try our best to make sure our words do not trigger her pain; if a family member asks for an asthma inhaler at the time we attend our friend's wedding, we will weigh their claims carefully and attempt to find alternatives that can satisfy both demands. Caution is the guiding virtue underlying the second principle, and I think it is also universally recognised that moral agents should be cautious.

These two principles can guide our judgment of whether a civil servant is culpable for her ignorance. The first principle entails that civil servants who follow state orders thoughtlessly are culpably ignorant. Saying 'I just followed orders without understanding them' will not be an excuse but rather a confession to prior failures of epistemic obligations. Unless the civil servant lacks the capacity for reflective self-control, playing dumb will only work against her. The second principle has more profound implications. The higher the rank of civil servants, the more essential functions they generally serve in the state machine that has pervasive influence over citizens' lives, and the higher the level of their epistemic obligations generally is. Also, the higher the rank they are, the more they already know about how their roles affect people's lives, and this generates a higher level of epistemic obligations too.

Similarly, those who directly execute or witness individual instances of state atrocities have more substantial epistemic obligations than others. It is because no civil servants who see the corpses, hear the screams, or witness the traumatic effects of sexual violence and forced separation can deny that what they do has colossal impacts on others' lives. Finally, members of the targeted population, some righteous-minded citizens, official and civil actors of foreign states, and international organisations can sometimes find opportunities to publicly reveal and criticise the occurring or imminent state atrocities. The louder and more visible their reveals are, the more their recipients know about the significance of what is the state is doing (they can recognise that there is an important issue even if they disagree with the criticisms), and so the more imperative it is for them to investigate the truth.

Whether a civil servant has fulfilled her respective level of epistemic obligations depends, ultimately, on her individual circumstances. Still, I think the above two principles suggest that the epistemic tasks ahead many of them cannot be not easily met. And unless they have taken their epistemic obligations seriously and made a substantial effort to fulfil them, they would remain culpably ignorant, and they could not claim a valid excuse of ignorance.

Epistemic obstacles

I must again note that we cannot directly determine the culpability of a civil servant's ignorance based on the content of their beliefs. This implies that despite the best epistemic efforts of a civil servant, she may remain mistaken about the nature of her actions, and her ignorance will remain non-culpable. This can happen for two reasons.

The first is our innate fallibility and limited capacity for observation and reasoning. Even scientists can often be wrong about facts. And as Harman (2011) notes, 'moral knowledge isn't easy to come by, and isn't assured if one manages one's beliefs well, not even if one is exposed to good moral arguments for the truth' (455). The second reason is that our capacity for acquiring factual and moral knowledge is partly dependent on the opportunities presented by our social circumstances. Sometimes, the social context may even present epistemic obstacles to our learning. I will discuss this reason in more detail.

To begin, some social and personal circumstances may prevent civil servants from knowing the existence of state atrocities and that their role contributes to the atrocities. Some of these obstacles include the distance between the civil servants' daily work and the atrocities, the secrecy of state wrongdoings, and the state's suppression of free information flow. As such, it may be impossible for some civil servants to understand the potential causal connection between their work and the atrocities. It may also be that their level of epistemic obligations is not high enough to overcome the obstacles, and even if the level of effort is high enough, some may still misunderstand the big picture. A fair judgment of the excuse of factual ignorance must take these possibilities seriously.

The literature has discussed many cultural obstacles to the acquisition of moral knowledge. For example, Isaacs (1997) observes that sometimes 'the issue that some minority of people recognize as morally charged is not recognized as a moral issue at all by the general population' (679). In other words, civil servants may not be aware of any moral controversies surrounding their action – even if they are ready to fulfil their epistemic obligations, they do not know which aspects of their actions to investigate. Furthermore, Benson (2001) points out that 'oppressive practices sustain themselves partly by systematically covering up evidence that the persons they oppress deserve better treatment' (611). He illustrates this point with Aristotle, who critically assessed and rejected the conventional justifications of slavery in ancient Greece, but still came to the wrong conclusion that slavery was good for those naturally born slaves. In this case, his mistake was not due to his inadequate epistemic effort, but rather that 'his aristocratic social circumstances may have occluded his access to good evidence about the rational capabilities of persons subjected to enslavement' (611). This observation suggests that, when atrocities happen in societies where the targeted group has long been denied opportunities to demonstrate their humanity, the perpetrators' limited access to evidence of the targeted group's humanity may justify their ignorance. Yet, where the targeted group has been neighbours and functioning members of society, it would be culpable to be ignorant of their humanity.

The epistemic value of state judgments may also present epistemic obstacles to civil servants' acquisition of moral knowledge. Epistemic value is the tendency of a judgment, factual or moral, to be right, grounded in the merit of the epistemic process by which the judgment is made. And when a judgment has high epistemic value, it is reasonable for others to trust the judgment, especially if they have not (or cannot) undergone an equally good epistemic process. For instance, we often trust the advice of medical experts more than ours regarding health issues. It is because the experts' advice is usually backed by research and long-standing knowledge about health. Furthermore, after reading a few news articles about a criminal suspect, we may have some idea about what she deserves. But we still tend to trust the court's verdict even

if it is different from our judgment, for we know that the judges have listened to and assessed the claims and evidence of both sides. I call this the practice of modest trust. This practice contains the virtue of modesty because it would be arrogant for a person to believe in the superiority of their opinions against others despite the absence of sound epistemic grounds. When climate change sceptics and anti-vaxxers are criticised, they are not criticised for not finding out the truth – ordinary people without scientific training cannot be expected to find out the truth anyway; instead, they are criticised for being arrogant – for not trusting the findings of the mass of research done with an extremely high epistemic standard.

The practice of modest trust opens two ways for civil servants to justify their ignorance in the context of atrocities. The first way is available when the state's judgment that it is right for civil servants to carry out its orders does have an epistemic value higher than their personal judgment. For example, the orders to commit atrocities may have resulted from a lengthy deliberative process about how to best serve the public good with inputs from experts and different groups. In these cases, even if the authorities' judgment is wrong, it has a higher tendency to be right than the private judgment of some civil servants. It would then be reasonable for some civil servants to trust the authorities' judgment that their contribution is made for the right cause. The practice of modest trust will justify their resulting ignorance. There is another way that the practice of modest trust may justify civil servants' ignorance even though the state's judgment lacks epistemic value. This is possible when the civil servant has non-culpably overestimated the epistemic value of state judgment, meaning that she has met her required level of epistemic obligations but still chooses to trust the state because she mistakenly believes that the state is more likely to be right than her. For example, she may have acquired from her upbringing a deep-seated belief that strong, adamant, masculine leaders are much wiser than ordinary people.

As such, due to the reasonableness of the practice of modest trust, the epistemic value of state judgments and the obstacles to gauging the epistemic value of state judgments both constitute obstacles for civil servants to understand the wrongness of their contribution. Here, some may doubt if we should really acknowledge the practice of modest trust because it seems to provide a blanket excuse of ignorance for all civil servants, and this implication may seem counterintuitive to some people. My response is that acknowledging the practice of modest trust does not create a blanket excuse of ignorance. It is because it is often quite easy for attentive citizens to tell that their government's judgments do not deserve trust in the context of atrocities. Whenever a government tells a blatant lie or shifts its pronounced moral judgment to secure its interests, it significantly damages the epistemic value of its later judgments, for honesty is also a significant determinant of epistemic value. If these occasions repeat and are visible to civil servants, they have a fragile claim that they still non-culpably believe in the epistemic authority of the state.

As Alvarez (1997) observes, some Nazi perpetrators 'saw their actions in the light of a hard but necessary patriotic duty. … it became a matter of the needs of Germany and the German people outweighing their internal prohibitions against killing and genocide' (165). Even if they committed atrocities only because they had trusted the authority of their leaders, the excuse of ignorance was not necessarily available to them. It was because, as Estlund (2007) suggests, 'Nazi soldiers and officers had no basis for thinking that the political source of their commands was duly looking after the

question of whether the war was just' (228). Their continuing trust towards the state was likely blind rather than modest, and ignorance due to blind trust is culpable.

To conclude, ignorance can be a valid excuse for civil servants, but it is valid only if the ignorance is not culpable. I agree with Zimmerman and Rosen that ignorance is culpable only if it is caused by a prior culpable failure of epistemic obligations. However, I disagree with them that culpable ignorance is rare because non-ignorant culpable acts must be akratic; instead, I argue that a choice can be culpable because the agent has a dispositional belief that the action is wrong, so culpable ignorance is not as rare as Zimmerman and Rosen suggest. I have also proposed two principles for guiding the determination of each civil servant's level of epistemic obligations. These principles show that thoughtless civil servants are culpably ignorant and that civil servants who are of higher rank, have closer contact with atrocities, and have more exposure to alternative views generally have a higher level of epistemic obligations than others. I also explain why some civil servants may remain mistaken about the nature of their role despite having spent adequate epistemic effort. In particular, I suggest that obstacles to gauging the epistemic value of state decisions may also prevent civil servants from recognising the wrongness of their contribution.

Conclusion

This chapter has examined four justifications and excuses commonly claimed by civil servants. The cog-theory justification is invalid because civil servants who have not caused state atrocities may still have contributed to them by risking their successes. The superior order justification is invalid because some civil servants have no contractual obligation to obey illegal superior orders, and even if they have not promised to obey the law and/or the atrocity-inducing orders are not illegal, they still have no obligation to obey these orders because bad promises do not obligate. The excuse of absent discriminatory intent is also invalid because a culpable choice involves a negative rather than a positive requirement. A choice is culpable if it is made with knowledge of the wrongness of the action, and this knowledge fails to demotivate the choice. However, I acknowledge the validity of the excuse of ignorance, but it is valid only if the ignorance is not culpable, meaning that the civil servant must have fulfilled her respective level of epistemic obligations.

Therefore, while it is wrong for civil servants to contribute to state atrocities, the conditions of punishing them justifiably are limited by the excuse of ignorance and other valid justifications and excuses. Again, the question of who should be punished depends ultimately on each civil servant's individual circumstances. But I believe the discussions of this chapter can still offer guidance to the determination of each civil servant's culpability.

Chapter 5
Conclusion

This thesis investigated whether civil servants be should for past state atrocities under transitional circumstances. Chapter 2 inquired into the sub-question of whether retributivism provides an application justification of punishment in transitional circumstances. Having addressed Murphy's challenges about the effectiveness of punishment, the standing of the punisher, and the retroactivity of punishment, I argued that retributive justice does offer a defeasible justification for punishing civil servants in transitional circumstances.

Then, building upon the principle of expressivist retributivism – culpable wrongdoers, and only culpable wrongdoers, deserve proportional punishment as blame – Chapter 3 looked into the sub-question of how civil servants can be held complicitly responsible for state atrocities. I rejected Kutz's, Lawson's, Gardner's, and Petersson's solutions to the problem of overdetermination, and then argued that an individual is complicitly responsible for a collective wrongdoing if she has taken an action that carries an impermissible risk of causing the collective wrongdoing's success, and this action expresses a culpable choice. This principle of complicit responsibility laid the foundation for holding civil servants responsible for atrocities.

Finally, Chapter 4 investigated the sub-question of what valid justifications and excuses civil servants can claim to counter blame and punishment. I surveyed the cog-theory justification, the superior order justification, the excuse of absent discriminatory intent, and the excuse of ignorance, and concluded that only the excuse of ignorance was valid. I also proposed how a civil servant's level of epistemic obligations could be determined and argued that obstacles to gauging the epistemic value of state decisions could prevent a civil servant from understanding the wrongful nature of their contribution to atrocities.

Therefore, my answer to this principal thesis question is this: A civil servant deserves to be punished for past state atrocities if she has contributed to them, in the sense of risking their successes, and she cannot appeal to any valid justification or excuse, such as the lesser evil justification, the excuse of duress, and the excuse of ignorance, to show that she is not culpable for her contribution. And if there are no forward-looking considerations, such as peace and political stability, weighty enough to outweigh the importance of achieving retributive justice in the transitional context, these civil servants complicitly responsible for past state atrocities should be punished.

At the beginning of this thesis, I pointed out that civil servants had rarely been the targets of punishment in post-atrocity trials. They usually enjoyed impunity for contributing to atrocities while the political leaders and high-level cadres were prosecuted and punished. Therefore, apart from contributing to existing discussions in the literature, I believe this thesis also shows that we should rethink our transitional justice practices. And if any civil servants contribute to state atrocities because they believe they should not be punished, they are terribly mistaken.

Bibliography

This bibliography includes all the works mentioned in this thesis.

Allais, L. (2011). Restorative justice, retributive justice, and the South African Truth and Reconciliation Commission. *Philosophy & Public Affairs, 39*(4), 331-363.

Alvarez, A. (1997). Adjusting to genocide: The techniques of neutralization and the Holocaust. *Social Science History, 21*(2), 139-178.

Arendt, H. (1994). *Eichmann in Jerusalem: A Report on the Banality of Evil.* Penguin Books.

Arendt, H. (2003). *Responsibility and Judgment*. Schocken Books.

Benson, P. (2001). Culture and responsibility: A reply to Moody- Adams. *Journal of Social Philosophy, 32*(4), 610-620.

Boot, E. R. (2019). Obligatory whistleblowing: Civil servants and the complicity-based obligation to disclose government wrongdoing. *Journal of Moral Philosophy, 16*(2), 131-159.

Crawford, K. L. (1990). Due obedience and the rights of victims: Argentina's transition to democracy. *Human Rights Quarterly, 12*, 17.

Dana, S. (2008). Beyond retroactivity to realizing justice: A theory on the principle of legality in international criminal law sentencing. *Journal of Criminal Law & Criminology, 99*, 857.

Drumbl, M. A. (2007). *Atrocity, Punishment, and International Law*. Cambridge University Press.

Drumbl, M. A. (2011). Collective responsibility and postconflict justice. In T. Issacs & R. Vernon (Eds.), *Accountability for Collective Wrongdoing* (pp. 23-60). Cambridge University Press.

Duff, R. A. (2001). *Punishment, Communication, and Community.* Oxford University Press.

Elster, J. (2004). *Closing the Books: Transitional Justice in Historical Perspective*. Cambridge University Press.

Erskine, T. (2011). Kicking bodies and damning souls: The danger of harming "innocent" individuals while punishing "delinquent" states. In T. Issacs & R. Vernon (Eds.), *Accountability for Collective Wrongdoing* (pp. 261-286). Cambridge University Press.

Estlund, D. (2007). On following orders in an unjust war. *Journal of Political Philosophy, 15*(2), 213-234.

European Union. (2015). *The EU's Policy Framework on Support to Transitional Justice.*

Fanon, F. (1961). *The Wretched of the Earth* (R. Philcox, Trans.). Grove Press.

Feinberg, J. (1970). *Doing & Deserving; Essays in the Theory of Responsibility*. Princeton University Press.

FitzPatrick, W. J. (2008). Moral responsibility and normative ignorance: Answering a new skeptical challenge. *Ethics, 118*(4), 589-613.

Galand, A. S. (2019). *UN Security Council Referrals to the International Criminal Court.* Brill | Nijhoff.

Gallant, K. S. (2009). *The Principle of Legality in International and Comparative Criminal Law.* Cambridge University Press.

Gardner, J. (2004). Book Reviews [Review of the book Complicity: Ethics and Law for a Collective Age, by C. Kutz]. *Ethics, 114*(4), 827-830.

Greenawalt, A. K. (2013). International criminal law for retributivists. *University of Pennsylvania Journal of International Law, 35*(4), 969-1044.

Hampton, J. (1992). Correction harms versus righting wrongs: The goal of retribution. *UCLA Law Review, 39*(6), 1659-1702.
Hansen, T. O. (2017). The time and space of transitional justice. In C. Lawther & L. Moffett (Eds.), *Research Handbook on Transitional Justice* (pp. 34-51). Edward Elgar Publishing.
Hansson, S. (2013). *The Ethics of Risk: Ethical Analysis in an Uncertain World*. Springer.
Harman, E. (2011). Does moral ignorance exculpate?. *Ratio, 24*(4), 443-468.
Insco, J. B. (2003). Defense of superior orders before military commissions. *Duke Journal of Comparative & International Law, 13*(2), 389-418.
International Criminal Court. (2011). *Rome Statute of the International Criminal Court.*
International Criminal Court. (2022). *The States Parties to the Rome Statute*. Retrieved from: https://asp.icc-cpi.int/en_menus/asp/states%20parties/Pages/the%20states%20parties%20to%20the%20rome%20statute.aspx
Isaacs, T. (1997). Cultural context and moral responsibility. *Ethics, 107*(4), 670-684.
Isaacs, T. (2011). *Moral Responsibility in Collective Contexts*. Oxford University Press.
Kalyk, M. (2018). *Crimes against Humanity in the 21st Century: Law, Practice and Threats to International Peace and Security*. Brill | Nijhoff.
Kleinig, J. (1973). *Punishment and Desert*. Martinus Nijhoff.
Kleinig, J. (2011). What does wrongdoing deserve? In M. Tonry (Ed.), *Retributivism Has a Past: Has It a Future?* (pp. 46-62). Oxford University Press.
Kumar, R. (2003). Who can be wronged? *Philosophy & Public Affairs, 31*(2), 99-118.
Kutz, C. (2000). *Complicity: Ethics and Law for a Collective Age*. Cambridge University Press.
Lawson, B. (2013). Individual complicity in collective wrongdoing. *Ethical Theory and Moral Practice, 16*(2), 227-243.
Lenta, P. (2019). Transitional justice and retributive justice. *Ethical Theory and Moral Practice, 22*(2), 385-398.
Lepora, C & Goodin, R. E. (2013). *On Complicity and Compromise*. Oxford University Press
Linton, S. (2001). Cambodia, East Timor and Sierra Leone: experiments in international justice. *Criminal Law Forum, 12*(2), 185-246.
Malamud-Goti, J. (1990). Transitional governments in the breach: Why punish state criminals. *Human Rights Quarterly, 12*(1), 1-16.
May, L. (2005). *Crimes Against Humanity: A Normative Account*. Cambridge University Press
McCoubrey, H. (2001). From Nuremberg to Rome: restoring the defence of superior orders. *International & Comparative Law Quarterly, 50*(2), 386-394.
McDermott, D. (2008). Analytical political philosophy. In D. Leopold & M. Stears (Eds.), *Political Theory: Methods and Approaches* (pp. 11-28). Oxford University Press.
McMahan, J. (2008). Collective crime and collective punishment. *Criminal Justice Ethics, 27*(1), 4-12.
Mokhtar, A. (2005). Nullum crimen, nulla poena sine lege: aspects and prospects. *Statute Law Review, 26*(1), 41-55.
Morris, C. W. (1991). Punishment and Loss of Moral Standing. *Canadian Journal of Philosophy, 21*(1), 53-77.

Murphy, C. (2017). *The Conceptual Foundations of Transitional Justice*. Cambridge University Press.
Nino, C. S. (1991). The duty to punish past abuses of human rights put into context: the case of Argentina. *The Yale Law Journal, 100*(8), 2619-2640.
Office of the United States Chief Counsel for Prosecution of Axis Criminality. (1946). *Nazi Conspiracy and Aggression Volume IV Document No. 2061-PS*. United States Government Printing Office. Retrieved from: https://avalon.law.yale.edu/imt/2061-ps.asp
Ohlin, J. D. (2009). Towards a unique theory of international criminal sentencing. *Cornell Law Faculty Publications*, 23.
Peels, R. (2011). Tracing culpable ignorance. *Logos & Episteme, 2*(4), 575-582.
Petersson, B. (2013). Co-responsibility and causal involvement. *Philosophia, 41*(3), 847-866.
Rosen, G. (2004). Skepticism about moral responsibility. *Philosophical Perspectives, 18*, 295-313.
Shahabuddeen, M. (2004). Does the principle of legality stand in the way of progressive development of law? *Journal of International Criminal Justice, 2*(4), 1007-1017.
Sikkink, K., & Walling, C. B. (2007). The impact of human rights trials in Latin America. *Journal of Peace Research, 44*(4), 427-445.
Simmons, A. J. (1976). Tacit consent and political obligation. *Philosophy & Public Affairs, 5*(3), 274-291.
Skerker, M. (2020). Individual responsibility for collective action. In S. Bazargan-Forward & D. Tollefsen (Eds.), *The Routledge Handbook of Collective Responsibility* (pp. 274-284). Routledge.
Stewart, J. G. (2012). Overdetermined atrocities. *Journal of International Criminal Justice, 10*(5), 1189-1218.
Stilz, A. (2011). Collective responsibility and the state. *Journal of Political Philosophy, 19*(2), 190-208.
The Crown Prosecution Service. (2019, February 4). *Secondary Liability: Charging Decisions on Principals and Accessories.* Retrieved from: https://www.cps.gov.uk/legal-guidance/secondary-liability-charging-decisions-principals-and-accessories
Thompson, D. F. (1980). Moral responsibility of public officials: The problem of many hands. *The American Political Science Review, 74*(4), 905-916.
Thompson, D. F. (1985). The possibility of administrative ethics. *Public Administration Review, 45*(5), 555-561.
Todd, P. (2019). A unified account of the moral standing to blame. *Noûs, 53*(2), 347-374.
Tutu, D. (2009). *No Future Without Forgiveness*. Doubleday.
United Nations Security Council. (2021). *International Tribunals*. Retrieved from: https://www.un.org/securitycouncil/content/repertoire/international-tribunals
United Nations. Secretary-General. (2004). *The Rule of Law and Transitional Justice in Conflict and Post-conflict Societies: Report of the Secretary-General*. United Nations Security Council.
Vernon, R. (2011). Punishing collectives: States or nations? In T. Issacs & R. Vernon (Eds.), *Accountability for Collective Wrongdoing* (pp. 287-306). Cambridge University Press.
Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*. Harvard University Press.

Weber, M. (2014). Politics as a vocation. In H. H. Gerth & C. W. Mills (Eds.), *From Max Weber: Essays in Sociology* (pp. 77-128). Routledge.
Wittmann, R. E. (2003). Indicting Auschwitz? The Paradox of the Frankfurt Auschwitz Trial. *German History, 21*(4), 505-532.
Zimmerman, M. J. (1997). Moral responsibility and ignorance. *Ethics, 107*(3), 410-426.