Causation, Responsibility, and Climate Change

A Thesis

submitted to

Indian Institute of Science Education and Research Pune in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Mohammad Wasim

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

April, 2023

Supervisor: Dr Tarun Menon

Name: Mohammad Wasim

All rights reserved

Certificate

This is to certify that this dissertation entitled 'Causation, Responsibility, and Climate Change', towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents the study/work carried out by Mohammad Wasim at Azim Premji University, Bengaluru, under the supervision of **Dr Tarun Menon**, Assistant Professor, School of Arts and Sciences, Azim Premji University, Bengaluru, during the academic year 2022-23.

Dr Tarun Menon

Committee:

Dr Tarun Menon

Dr Pushkar Sohoni

This thesis is dedicated to my family

Declaration

I hereby declare that the matter embodied in the report entitled 'Causation, Responsibility, and Climate Change' are the results of the work carried out by me at the School of Arts and Sciences, Azim Premji University, Bengaluru, under the supervision of **Dr Tarun Menon** and the same has not been submitted elsewhere for any other degree

Mohammad Wasim

Date: 01/04/2023

Table of Contents

Abstract

This thesis presents a framework for understanding moral responsibility in the context of climate change. After reviewing existing literature on responsibility, the author argues for a forward-looking approach that takes collectives as primary bearers of responsibility. This framework is grounded in the interventionist framework of causation, which allows for a better understanding of the ways in which our actions can intervene in the complex causal relationships underlying climate change. The author proposes a new kind of agency, known as an intervention agency, that is necessary for the ascription of intervention responsibility. Additionally, the author provides an account of intervention strength using resources from the interventionist theory of causation and causal emergence. The author also argues that coarse-grained modelling can provide a useful tool for understanding the ways in which collective agents have more capacity to intervene in these causal relationships than individual agents. The author emphasizes the importance of different kinds of justice as generators of responsibility, rather than as grounds. The thesis concludes by acknowledging that while the framework provided in this thesis is useful for understanding moral responsibility, further work is needed to formalize and evaluate the proposed notion of responsibility more precisely, as well as to address the normative aspect of intervention responsibility. The author hopes that this thesis will contribute to the growing field of climate ethics and encourage recognition of our individual and collective responsibilities to mitigate the effects of climate change.

Acknowledgments

I would like to express my heartfelt gratitude to my family, whose unwavering support and encouragement have been the cornerstone of my life. They have been a constant source of inspiration and strength, and without their love and support, this thesis would not have been possible.

I would like to extend my sincere thanks to my supervisor, Dr Tarun Menon, who has been a constant source of guidance, support and encouragement since the time I started working under him. His friendly nature, academic expertise and the freedom he has given me to explore my ideas and research have been invaluable, and I am extremely grateful for his mentorship. I would also like to thank Dr Pushkar Sohoni for graciously agreeing to serve as the expert on my thesis.

Finally, I would like to express my deep appreciation to my sponsor, Mr Sonam Wangchuk, the founder of the Himalayan Institute of Alternatives Ladakh, whose generous grant has provided me with the financial support necessary to pursue this project without any financial burden. His support has been instrumental in enabling me to devote my time and energy to this research, and I am grateful for his faith in my abilities.

Once again, I extend my heartfelt thanks to my family, supervisor, and sponsor for their support and encouragement, which has been invaluable to me throughout this project.

Chapter 1
Introduction

The problem of climate change is a pressing and urgent issue that has garnered widespread attention in recent years. The scientific consensus is clear: human activities, particularly the burning of fossil fuels, are causing unprecedented changes in the earth's climate, resulting in rising temperatures, melting glaciers, and more frequent extreme weather events.

The consequences of climate change are far-reaching and potentially catastrophic, affecting not only natural systems but also humans who are living now as well as those yet to come into the world. Climate change threatens global Justice; the poorest and most vulnerable populations, particularly those in developing countries, are likely to bear the brunt of the impacts of climate change, despite being least responsible for its causes. Climate change poses a significant threat to intergenerational justice, as it has the potential to cause irreversible harm to future generations. Furthermore, it threatens ecological justice by disrupting natural systems and ecosystems, which undermines the ability of all species, including humans, to thrive (Gardiner 2011). Given the magnitude and urgency of the problem of climate change, there is a growing need for ethical considerations and moral responsibility in addressing it. Climate change is not merely a scientific, economic or political issue, but a moral one, as it raises fundamental questions about our responsibilities to future generations, the global poor, and the planet as a whole (Jamieson, 2014, pp. 167-201).

The issue of moral responsibility for climate change is complex and multifaceted, and it raises important questions about the nature of individual and collective responsibility as well as forward and backward-looking notions of responsibility. Some argue that individual actions, such as reducing one's carbon footprint or using public transport, are the key to mitigating

climate change (Jamieson, 2007). However, such an approach fails to take into account the systemic nature of the problem and the fact that climate change is the result of the cumulative actions of individuals, corporations, and governments over decades. Others argue that collective responsibility is the key to addressing climate change. In this view, responsibility for climate change is primarily borne by corporations, nations, intergovernmental organisations, etc. (Sinnott-Armstrong, 2010). Along with this, the responsibility for climate change is usually stated in backward-looking way, formulated in terms of historical responsibility and blameworthiness. However, some recent philosophers have argued that we must adopt a forward-looking approach to responsibility.

Against this backdrop, a consequentialist approach to moral responsibility can provide us with a useful framework for addressing the problem of climate change. This approach, more specifically understood in the literature as forward-looking responsibility, focuses on the outcomes of our actions and their impact on the well-being of all living beings, including future generations. By adopting a consequentialist framework, we can gauge the responsibility of different agents based on the feasibility and effectiveness of their actions in bringing the desired outcomes. Based on the broad principle of moral consequentialism (Sinnott-Armstrong, 2019), I develop a forward-looking model of responsibility, called intervention responsibility that primarily sees collectives and group agents as the primary bearers of this responsibility. The proposal of a forward-looking collective responsibility is motivated by the fact that only collective agents such as nations, corporations, and intergovernmental organisations are best suited to tackle the problem of climate change. Furthermore, I take a functionalist stance on the nature of group agency that is required for the attribution of responsibility to groups. As a response to the sceptics of group agency, I take the position that groups can have a kind of agency that is distinct from the agency of individual members, and

that this agency arises from the group's ability to function as a single entity in some crucial respects.

The primary task of my thesis, after reviewing the existing literature and taking stances on particular philosophical issues, is to provide a rigorous grounding for responsibility in the context of climate change. This task requires an understanding of the complexity and scale of the climate change problem, as well as a clear and coherent framework for assessing moral responsibility. To this end, I draw on the interventionist framework of causation to understand the forward-looking notion of responsibility. This approach is novel and original and allows for a more nuanced understanding of the complex causal relationships that underlie climate change. By understanding the ways in which our actions can intervene in these causal relationships, we can better understand our individual and collective responsibilities to mitigate the effects of climate change. In addition to the interventionist framework of causation, my thesis also draws on the coarse-grained modelling framework to understand the ways in which collective agents can intervene in the causal relationships underlying climate change. I argue that coarse-grained agents, which are essentially collective agents, are more effective in intervening in these causal relationships, and thus bear a greater responsibility for mitigating the effects of climate change.

Notions of responsibility are constructed and utilized for normative purposes and they play vital functional roles in the overall machinery of our society. If one notion of responsibility seems out of place for a particular problem, it should be revised or new notions must be

constructed that are suitable for the issue at hand (Jamieson, 2015). The revision, construction, implementation, or evaluation of such notions is known in the literature as conceptual engineering (Chalmers, 2020). In this thesis, I use the method of conceptual

engineering to explicate the model of intervention responsibility and I aim to develop a clear and coherent account of responsibility that is useful for addressing the ethical problems that arise in climate change.

In the second chapter, I provide a thorough literature review of the existing theories of moral responsibility and collective responsibility, highlighting their limitations and challenges in the context of climate change. I argue that our existing models of forward-looking responsibility fail to capture the complexity of the climate change problem and suggest a novel forward-looking responsibility that provides a thorough account of the capacity of intervention agents.

In the third chapter, I examine the concept of intervention and the responsibility to intervene, drawing on applied ethics to identify some of the problems associated with these two ideas. I then suggest a formalized understanding of intervention that can help us better understand its most general features.

In the fourth chapter, I introduce causal modelling and the interventionist framework of causation. I argue that an agent is an entity that has the tendency to represent causal interventions and causal models. Furthermore, I contend that we need to include the intervention agents inside the system and treat them as variables that have the tendency to intervene with other variables, thereby providing an objective way to assess various agents' strengths using causal models that represent the system.

In the fifth chapter, I define intervention responsibility in terms of intervention agency and intervention strength. I argue that agents with more intervention strength and agency are better equipped to intervene on other variables, and as such, bear more responsibility for bringing about certain outcomes, that are desirable, through their interventions.

Furthermore, I briefly suggest a way to incorporate historical responsibility for climate change in my account of intervention responsibility.

In the sixth chapter, I use coarse-grained modelling to argue that the system, which includes both agential and non-agential variables, must be viewed using a macro model that represents the interaction of coarse-grained agents with other variables. I suggest that coarse-grained agents, which are essentially collective and group agents, have more intervention strength and agency to intervene on other variables and thus bear heavier intervention responsibility. I argue that this understanding can help us better understand how to allocate responsibility in a collective action problem like climate change.

In the final chapter, I examine the social ontology of coarse-grained agents and suggest various criteria that must be kept in mind while coarse-graining individual agents.

Chapter 2
Literature Review

Moral responsibility is a widely used notion that pervades our personal lives as well as social, legal, and political domains. In philosophy, this concept has received considerable attention in the field of ethics. When we say that a person is morally responsible for a particular state of affairs, two implicit assumptions underlie this statement. First, there is a certain conception of moral responsibility that defines the term. Second, there is a model of responsibility that sets the conditions under which someone is held responsible.

To a first approximation, moral responsibility is generally understood to be a moral status ascribed to an agent for a particular state of affairs, when that agent possesses certain powers and capacities, and the state of affairs arises from the exercise of these powers and capacities (Talbert, 2022). The ascription of this moral status is then understood to elicit certain responses such as blame or praise. The effort to characterize these powers and capacities and their relationship to the state of affairs has been a major part of the debate on this topic for most of the last century.

For a long time, the power mentioned above was thought to be an agent's free will, and the debate on moral responsibility revolved around free will, mostly in the context of the compatibilism-incompatibilism debate (Talbert, 2022). This debate concerns the compatibility and incompatibility of free will with the causally deterministic universe. In these debates, compatibilism argued that responsibility was attributed to agents who caused a particular state of affairs through free action, understood as the exercise of free will. However, this foundational debate did not lead anywhere, at least in applied fields. By the latter part of the

last century, researchers started paying attention to a looser relative of free will, since free will is extremely hard to pin down. This notion is known as intentionality, which is generally understood to be the capacity of minds to be about, represent, and stand for things, properties, and states of affairs (Pierre, 2023).

Theories of moral responsibility are concerned with providing the correct model of responsibility and setting out the grounding conditions that explain the ascription of responsibility to an agent. Traditional models of moral responsibility ground moral responsibility in the agency of the agent, which is the capacity of an agent to act in certain ways (Schlosser, 2019). The literature on the philosophy of action provides a conception of agency, which explains agency in terms of intentionality. In this light, an agent is held responsible if it caused a state of affairs through the exercise of its agency, explained further in terms of intentional action. That is, the agent must have caused it through intentional action. Another important ground of moral responsibility is causation. An agent is held responsible if it caused the state of affairs (Talbert, 2022).

Problems of traditional accounts of moral responsibility

The traditional account takes agency, understood narrowly in terms of intentionality, as a necessary condition for moral responsibility on the one hand, and causation, an unnecessary condition on the other hand. However, it has been argued that the presence of intentionality, as well as causation, is insufficient to account for moral responsibility. A person who lacks the capacity to understand the moral consequence of her action does not seem to be an appropriate site of responsibility, even if she intended to cause a particular state of affairs

(Talbert, 2022). For instance, a child who breaks a stranger's laptop with the intention to break it lacks the ability to understand that it is morally wrong to do so. Or a person who is morally impaired and can't understand morality at all and causes a particularly undesirable state of affairs can be exempt from ascribing moral responsibility. This ability to recognize and respond to moral demands is known as moral competence and it is a constraint on moral responsibility. Another constraint is the ability to foresee the consequences of the action (Talbert, 2022). If a person is ignorant of the consequences, then, in some cases, the agent can be excused. However, it can only be excused for ignorance if that ignorance itself issues from an act for which it is not culpable. For example, if a newly graduated medical surgeon performs a surgery, which results in some harm because she missed an essential step in the process of surgery due to her ignorance of the fact that that step was essential-which further stemmed from another fact that she was not taught about that step by her professor-then she can be excused. This condition on moral responsibility is called epistemic competence and it undermines the traditional account of moral responsibility grounded in intentionality. While these two conditions definitely point out the insufficiency of intentional agency as a ground of moral responsibility, they do not suggest that intentionality is unnecessary.

Insofar as causation as a ground of moral responsibility is concerned, the problems are far graver. Some have argued that causation alone is insufficient, while some even go further to argue that causation is unnecessary (Sartorio, 2004), to ground moral responsibility. The examples given above serve as illustrations of this argument too. A person without moral or epistemic competence can be exempt from attributing moral responsibility even if she is the cause of a particular undesirable state of affairs. While the insufficiency of causation as a ground seems fairly uncontroversial, the claim about its unnecessity is quite counterintuitive.

It has been argued that moral responsibility obtains without causation; a person can be held responsible for a state of affairs without even causing it (Sartorio, 2004).

Another influential way to look at moral responsibility is to distinguish between forward-looking and backward-looking notions of moral responsibility (Talbert, 2022). Traditional models of moral responsibility, as we discussed, have mainly taken the backward-looking approach to moral responsibility. This notion of responsibility seeks to blame a person for having caused a particular state of affairs by looking at past actions that lead to the state of affairs. It is also known as retrospective responsibility and it seeks to assign responsibility based on actual causation. This approach emphasizes the role of causation in determining responsibility and places the burden of responsibility on the individual that caused the negative outcome. The forward approach to responsibility, also known as prospective responsibility, emphasises the advantageous outcomes of the practice of responsibility ascription, in contrast to the backward-looking responsibility which emphasises delivering punishment for past actions as a way to serve justice. These two notions of responsibility stem from different perspectives of justice, viz, retrospective and prospective justice. The forward-looking approach, although had fallen out of fashion in the last century, has found renewed interest among scholars in recent times. Another formulation of forward-looking responsibility focuses on the future consequences of actions and seeks to assign responsibility based on anticipated outcomes. In this latter formulation of forward-looking responsibility, the responsibility falls not on the agent who was complicit in bringing about a negative state of affairs, but rather on the agent who can mitigate the problem. While the notions of responsibility discussed so far assumes an individual agent as foci of responsibility ascription, it misses an important kind of agent, known as collective agents that is common in our practices of responsibility ascription.

Collective responsibility

The notion of collective responsibility refers in the most context to both the causal responsibility of agents for the harm in the world and the blameworthiness that we ascribe to them for having caused such damage through their collective action (Smiley, 2022). Similar to its individualistic counterpart, collective responsibility is analogously grounded partly in the group agency of the collective, and partly in group-causation. Group agency (Roth, 2017) is the tendency of a group to act by deliberation, use of reason and the ability to intend as a collective. Similar to individual agency, group agency is also usually understood in terms of the collective intentionality of the group. However, recent scholarships have revealed that there is a multitude of features and conditions that ground group agency. Before we go on the discuss all the grounding conditions, a discussion of group agency is worthwhile.

The greatest criticism of collective responsibility comes from individualists who deny the legitimacy of group agency as well as the normative merits of collective responsibility (Smiley, 2022). Methodological individualists dispute the possibility that groups can have agency, which is distinct from the summation of the individual agencies and normative individualists fear that collective responsibility undermines the ideals of individual liberty and fairness. The debate can be traced back to the 19th-century debate between individualism and holism. In recent times, a new surge of arguments has emerged in defence of group agency as a legitimate property borne by groups (Epstein B. , 2017) (Gilbert M. , 1989) (Ritchie, 2015). These efforts have attempted to defend the view that group agency is a distinct property that is greater than the sum of individual agencies that constitute it. Although group agency is understood to supervene on individual agencies, it can't be reduced to individual agencies since there are other things on which it supervenes too (Epstein B. , 2015). These contemporary

debates on group agency started from John Searle's work on collective intentionality (Smiley, 2022), which sought to explain group-level phenomena in terms of the collective intention of a group and has matured over the decades. Margaret Gilbert (Gilbert, 1989) has argued that group agency can be understood in terms of the joint commitment of its members whose wills get forged into a unified will. Michael Bratman (Bratman, 2014) analyses group agency in terms of the shared intention of its members. Some have argued that group agency arises due to the structure of the group in which members are arranged (Ritchie 2015), while some have endorsed heterogeneity of conditions that give rise to group agency (Epstein 2015, 2019). The structural condition of group agency is important because not every random collection of agents have the ability to commit and intend jointly. It is only certain groups having particular structures that have the kind of agency needed for a responsibility ascription. In particular, the groups must have a well-organised body, consisting of a representative governing body that represents the intention and beliefs of the group. Peter French considers three salient features for groups to be appropriate sites of collective responsibility; "

1. Series of organizational mechanisms through which courses of concerted action can be chosen and taken.
2. Set of enforced standards of conduct for individuals of the group.
3. a configuration of "defined roles" by which individuals can exercise certain powers .." (French, 1984, pp. 13–14)

The notion of collective responsibility we have discussed thus far has been retrospective responsibility or backward-looking responsibility which is ascribed to a group for having caused particular harm in the world. However, there are many cases where the ascription of backwards-looking is problematic as well as futile. In cases of structural injustices and

tragedies of the commons, it is not possible to hold someone responsible for causing harm because the contribution is extremely minuscule. Additionally, it is not appropriate to hold individuals responsible to remedy the harm either, because the individual's efforts will be again insignificant. In such cases, the responsibility to remedy the harm is usually assigned to group agents and collectives and such responsibility is a prospective or remedial or forward-looking responsibility. Forward-looking collective responsibility (FLCR) takes responsibility to be a matter of morally charging agents to bring about certain states of affairs, usually mitigating harm, that we as a community deem as undesirable (Smiley, 2022). Like its individualistic counterpart, FLCR has two formulations. On one formulation, it is assigned to the collective which was causally complicit in producing the morally problematic state of affairs and the purpose of the assignment is to mitigate the harm and bring about desirable changes in the existing state of affairs. To this end, some ground forward-looking collective responsibility in causal responsibility of the group. David Lyons (Lyons 2003) has argued, with regard to poverty and racism in the US, that the US was causally responsible for this state of affairs through its policies and decisions in the past, and now it has the responsibility to remedy the situation. Others have argued that FLCR can't always be grounded in causal responsibility since the parties involved in producing the harm might not be in the best position to remedy the harm. Furthermore, the FLCR in such cases should rather be assigned to parties that have the best chances of remedying the harm. This strategy has been criticised for exempting the parties that were involved in producing the harm from bearing any responsibility, thereby violating the idea of fairness. Still others have tried to incorporate causal responsibility in FLCR without essentially grounding FLCR in causal responsibility. Tracy Isaacs (Isaacs, 2014) has argued that those who are causally implicated in producing the harm and benefit from it bear a heavier responsibility to mitigate the harm, thereby suggesting that

causal responsibility plays a role in identifying the agent that is to be assigned remedial responsibility. In the case of the tragedy of commons, generic instances of which are overgrazing, deforestation, global warming, etc, FLCR has been argued as the only pragmatic way of assigning responsibility, simply because other approaches are not feasible at all. The individual agents in the case of the tragedy of commons, who are otherwise disunited and working on self-interests and do not have the capacity to make a meaningful contribution individually, have the responsibility to come together, join hands, and mobilize themselves into organised groups, which can then deliver feasible actions to mitigate the harm collectively (Collins, 2012) (Hindriks, 2019). This process of joining hands gives rise to a certain kind of group agency in the collection of people, thus transforming a collection of people into a collective agent, which then becomes a legitimate site of FLCR.

So far, I have reviewed the literature on moral responsibility, its grounding conditions and problems, followed by the distinction between individual and collective responsibility. I have then reviewed the literature on collective responsibility and its grounding conditions, followed by its problems and merits. I have also discussed the distinction between forward-looking and backward-looking approaches to responsibility, which are crucial in the contemporary debates on responsibility, both in the case of individual responsibility as well as collective responsibility. In the next section, I will zero in on the specific problems that our traditional models of responsibility face in the context of climate change.

Climate change and the problems of traditional models of responsibility

Some describe climate change as a perfect moral storm (Gardiner, 2011), a novel threat to our common-sense morality. It defies all the moral intuitions that we have developed so far

(Gilbert D. , 2006). Mainly because the scale and the complexity of the crises are unprecedented. It is unlike any other phenomenon that demands our moral judgment. It is argued that an important thread in the tapestry of our morality is harm causation, where the cause, effect and causal chain obtain within a confined spatiotemporal interval. However, climate change is an immensely complex issue for which not only our ordinary moral intuitions, but most of our moral theories fall flat. Our moral framework is not sophisticated enough to account for the various calamities that climate change is expected to bring. In this section, my aim isn't, however, to examine the failure of our entire moral framework but rather to look at the problems that our traditional accounts of responsibility face in the climate crises. The responsibility to mitigate greenhouse gas (GHG) emissions is generally understood in terms of historical responsibility, where certain communities and countries are said to have contributed far more than others to the net increment in atmospheric GHG historically, and those countries and communities are therefore, said to be more blameworthy (Jamieson, 2014). And it is in virtue of this blameworthiness those countries are morally responsible to take climate action now. However, as I have discussed before, backward-looking responsibility grounded on blameworthiness faces several problems generally. And those issues crop up in the case of historical responsibility for climate crises too. It is argued that those countries which have historically contributed can't be blameworthy because they did not have knowledge of the harmful consequences of GHG emissions. The harmful consequences of GHG emissions were neither foreseen nor intended by the parties which were complicit. Another problem it faces is with regard to causal responsibility assigned to them for the emissions that they have historically produced. The causal relation between particular acts of emissions and particular instances of warming is impossible to be established. Causal responsibility is generally assigned when a clear causal link between two events is established

and the events, known as causal relata, are identified. However, it is epistemically impossible to establish such a relationship between a particular agent's emissions and particular warming (Sinnott-Armstrong, 2010). Even if we consider responsibility for emissions in present times, it is not possible to attribute specific emissions as the cause of certain effects observed in terms of drought, extreme weather etc. An emission from the US can have its effect in Africa and an emission produced in 2023 can have its year 3000, or maybe longer. There is no spatial or temporal contiguity between the cause and the effect which makes it impossible to attribute certain parties to causal responsibility. This is a serious problem for both global justice and intergenerational justice. Studies have shown that most of the emissions come from only certain countries (Jones et al. 2023), the majority of which are developed economies that are in the northern hemisphere, whereas the effects have been observed in the countries that lie in the southern hemisphere, most of which are developing nations. The idea of global justice demands that those developed nations emitting way more GHG than the developing nations that are at the receiving end of this calamity be made to pay in some way. But as I showed, the ascription of a backward-looking responsibility like historical responsibility based on causal responsibility will be problematic because the causal relations that are needed for the ascription of a backward-looking responsibility, are not possible to be established. There is really no way to trace the causal history of a particular effect at one geographical location and attribute a specific emission as its cause. Similarly, intergenerational justice demands that those living now be obligated to live in certain ways so that the people who are yet to come into the world have access to basic necessities and are able to live decent lives. But obligation generated by intergenerational justice can't be fulfilled by a backwards-looking notion of responsibility because the effect has not yet been realised. There is no law that guarantees that my particular emission will definitely contribute to the

warming in future. Due to these problems, moral responsibility understood as grounded in blameworthiness and causal contribution is not at all suitable to address the issues of global and intergenerational justice.

Climate change is a collective action problem and has often been categorised into the family of moral dilemmas known as the tragedy of the commons. I explained briefly why it is labelled as an instance of this problem. However, I must tell you that climate change is considerably different from other cases of this dilemma in many important respects. There are certain features of climate crises that set it apart from the normal instances of the tragedy of the commons, making it one of the most complex collective action problems. The common model (diagram given below) that is used to illustrate collective action problems like overgrazing involves a system with a large number of causal contributions each having an associated effect. Although, both the contributions and the effects might be practically imperceptible, in principle, every input must have an associated output in this model. And a large number of imperceptible inputs give rise to a significant output. Although this model might seem appealing for its simplicity, it does not capture the various complexities that are involved in climate crises. First of all, not every emission is imperceptible as it is generally considered in the tragedy of commons. Emissions from some mega-corporations do have significant impacts on the overall GHG concentration, at least their emissions averaged over a period of time. But even if we consider the system to be constituted of only individuals emitting imperceptible amount of GHG, their causal contributions and the relevant effects are not as straightforward as this model suggest. This common model, known as the Cumulative model (Jamieson, 2015), assumes that every input produces an effect in a causally robust manner. But the atmospheric chemistry of GHG is far more complex than this model seems to suggest. Between emissions and the final effect, there are various feedback loops and mechanisms that take place in the

biosphere and the atmosphere (Sinnott-Armstrong, 2010). A GHG emitted by a person driving his car might get absorbed by the trees around or a water body, where the GHG molecules might stay for centuries. The fate of the emitted GHG molecules is immensely unpredictable considering the non-linearities and complexity of the climate system. It is epistemically impossible to trace the causal chain from a particular emission leading up to a particular effect (Jamieson, 2015). In contrast, in the case of overgrazing, the causal relation between a single cause (animals eating the pasture) and its effect (pasture's grass getting depleted) are robust and deterministic, however imperceptible they are. I had discussed that it is not considered appropriate to hold individuals responsible to act in cases of the tragedy of commons because their individual contributions are not significant to have a meaningful impact. This feature of climate change we just discussed further undermines individual responsibility because it is not even clear whether individuals are really involved in contributing to the problem in the first place. This has motivated some people like Walter Sinnott-Armstrong to claim (Sinnott-Armstrong, 2010) that 'my individual joy ride does not cause global warming, climate change, or any of the resulting harms, at least directly'. But what the responsibility of individuals amounts to, he further claims, is to hold the govt responsible for making it illegal for him to joy-ride.

It is clear that our traditional accounts of responsibility grounded in blameworthiness and caution are ill-equipped to account for a problem of this magnitude and complexity. And this is perhaps reflected in our ordinary moral judgments. Climate change does not make our blood boil, the way the sight of a murder or an instance of blasphemy or even cases of gay sex disgust certain people, because our ordinary moral tapestry does not have the sophistication to make moral evaluations in climate change. It is been suggested that there should be a revision in our models of responsibility a way that the parties that are involved in contributing

to the problem are fairly obligated to take their share of action against the crises. Models of responsibility are constructed and utilised for specific purposes (Jamieson, 2015) and if our models of responsibility are inadequate to account for a particular case, we must strive to build new models. To this end, Dale Jamieson has proposed a notion of responsibility that is specially designed to be applied in the case of climate change. It is formulated as follows;

Agent A is IR for state of affairs S when

1) S is undesirable,

2) A could significantly mitigate S without excessive cost. (Jamieson, 2015)

This notion of responsibility is a forward-looking responsibility that appeals to the capacity of the agent. So, like other notions of forward-looking responsibility that have been proposed in various other cases, intervention responsibility is a responsibility that is grounded on the capacity of the agent to mitigate. While the name and the formulation of the notion of this responsibility sound quite apt to me, it lacks elaborate explication and rigorous grounding. There are a few questions that are left unanswered, some of which Jamieson acknowledges. The first question pertains to the criteria that make an entity an agent, as to what counts as an agent. The second question is regarding the features that determine the capacity. This question about the correct account of capacity is quite important and can be raised against other accounts of forward-looking responsibility that appeal to capacities too. It is usually argued that an agent is responsible to remedy harm if it has the 'capacity' to mitigate the harm. Additionally, while the forward-looking models of responsibility are generally proposed to be not grounded causation, in contrast to the backward-looking models, some notions that are essential to forward-looking responsibility are very causal. For instance, the notions of mitigation, intervention, influence, and production that the proposed idea of capacity is

supposed to deliver, are all causal. Although all the traditional models of forward-looking responsibility give a central role to the notion of influence, they do not give any elaborate account of the nature of this notion. Just as backward-looking responsibility was grounded on actual causation (contributory causes if there are multiple causes), a salient feature, if we want to formulate forward-looking responsibility that is grounded on the capacity to influence and explain the causal nature of this influence, we face a challenge: what kind of causal notion must we adopt to provide an adequate account of the nature of this influence?

I argue that not only we must take a forward-looking approach to responsibility in the case of climate change but also revise and extend the existing forward-looking models of responsibility that are commonly used in other collective action problems. There are a couple of reasons for this. Climate change is substantially different from other collective action problems, owing to its scale and complexity (Jamieson, 2015), so we must develop a more sophisticated model that adequately accommodates the nuances that emerge in climate crises. Furthermore, the traditional models of forward-looking responsibilities have not provided any elaborate account of certain notions that are so central to the models, such as agent, influence, and intervention. In the rest of this thesis, I aim to develop this notion of intervention responsibility, by making it more nuanced and giving a rigorous ground for the various notions associated with it. I aim to provide a full-blown account of intervention responsibility by illuminating the nature of agents, intervention and the capacity.

Chapter 3
Understanding Intervention and Responsibility

Intervention is generally understood to be a kind of interference in a process or condition, resulting in a change in the outcome of the process or a difference in the state. Colloquially, it refers to the process of interfering with the outcome or course of a condition or process, mainly to prevent harm or improve functioning. Examples of interventions include educational interventions that aim to provide support to help students acquire certain essential skills and surgical or medical interventions that aim to prevent various illnesses. Intervention can also be an interference by one state in another state's affairs with the aim of forcing it to do or refrain from carrying out certain acts. Another way to think of intervention is as an act of becoming intentionally involved in a difficult situation to improve it or prevent it from getting worse. For instance, it could be an occasion when someone's friends or family speak to them about a problem or situation because the person's behaviour is unreasonable or harmful.

Broadly construed, intervention is a means-to-end act on a state of affairs, where the end result aims to be positive or better in some sense, and the existing state of affairs is undesirable. The state of affairs may involve physical states, biological states, humans, social groups such as nation-states, classes of people, etc. The intervener is usually an agent, whatever the requirement of the agent is taken to be. Another feature of intervention is intentionality, as the intervention agent is intentionally involved in the act. The agent is able to form intentions, whatever the correct account of intentionality is. Therefore, it is not a random process that interferes with a state of affairs, but an intentional act from an agent.

The existing state of affairs is undesirable in the sense that it is abnormal, contrary to our interests, harmful, or can potentially lead to harm in the future, and are in need of corrective or preventive measures. The state of affairs may involve a physical state, such as a transformer malfunctioning that might lead to a power failure in the future, or a biological state, such as an organ malfunctioning that might lead to a life-threatening condition. It may also be about the psychology of a person, such as someone suffering from mental health issues and in need of counselling and other measures, or a group of people that needs aid for various reasons. Some of the kinds of interventions that are often talked about in public are military intervention, humanitarian intervention, medical or surgical intervention, and divine intervention. Intervention can also be appropriate in less disastrous cases, such as the act of intervention in a brawl with the intention of stopping it from getting worse.

Up to this point, we have discussed the act of intervention in various scenarios without delving much into the conditions and rules that necessitate intentional interventions. Questions that target these issues are: why must an agent intervene in a state of affairs that is undesirable? What duty does the agent owe? Does a person have a certain duty to end a brawl? Does a state have a duty to intervene in another sovereign state's affairs if that state is violating certain widely accepted norms? What kind of duty, if at all, does a physician have towards an injured person?

In various cases, agents are taken to owe certain duties or responsibilities. The nature of these responsibilities will be the interest of our discussion henceforth. I will explore the literature on the notions like intervention, responsibility, and their interrelations, in applied ethics and related areas, to get a sense of the technical usage of these notions in these areas. Some particular features of this responsibility that I will discuss are the conditions on which it depends, what explains the existence of this construct, etc. In technical terms, we will be

specifically interested in understanding the grounding schemes of intervention responsibility. The literature on topics related to intervention and the responsibility that is invoked in other areas of applied ethics is rich and can be potentially fruitful to examine for our interest.

Humanitarian Intervention

The act of using military force by a state in another state's territory to stop it from violating various human rights without the permission of the host state is generally referred to as humanitarian intervention. It has been a practice for many centuries in the international political arena. There have been various occasions in history where interventions have been carried out for supposed violations of the rights of certain ethnic groups, such as western intervention in the Ottoman Empire. However, after the Second World War, with the formation of the UN, new laws about sovereignty were formulated, and military intervention was largely considered an illegitimate action.

Humanitarian intervention is understood to be justified by intervening states as having some kind of right to intervene in certain situations. Various people have defended humanitarian intervention on different grounds. J.S. Mill wrote in his 1859 essay, 'A Few Words on Non-Intervention' (Mill, 2020):

> "There seems to be little need that the whole doctrine of non-interference with foreign nations should be reconsidered, if it can be said to have been considered as a moral question at all... To go to war for an idea, if the war is aggressive, not defensive, is as criminal as to go to war for territory or revenue; for it is as little justifiable to force our ideas on other people as to compel them to submit to our will in any other respect. But there are cases in which it is allowable to go to war,

> without having been ourselves attacked or threatened with attack; and it is very important that nations should make up their minds in time as to what these cases are... To suppose that the same international customs and the same rules of international morality can obtain between one civilized nation and another, and between civilized nations and barbarians, is a grave error...."

He justifies his conclusion that interventions in barbarian lands are sometimes permissible by the virtue of them not being 'civilized' enough. And the intervening civilized state has a moral right to intervene in their land and make them civilized. His conclusion, of course, comes from his clear imperialism. There is no legitimate ground for his argument.

John Rawls (Rawls,1993) gave his theory of humanitarian intervention based on what he calls a 'well-ordered society'. A well-ordered society is peaceful and respects the basic human rights of all its citizens. Between such societies, the principle of non-intervention must be upheld. However, societies that are not well-ordered - which are oppressive to their citizens or expansionist in attitude - must not be shielded from the principles of non-intervention. In such cases, states have the right to intervene. Martha Nussbaum (Nussbaum, 2005) criticizes Rawls's account for being too statist. It is concerned with whether certain states are well-ordered or not, instead of bringing the suffering of individuals to the forefront. The sufferings of people are the foundation of humanitarian intervention rather than impersonal states. She proposes the idea of human capability and argues that sovereignty should be respected, within the constraints of promoting human capabilities such as the capability to live a healthy life. If a state fails to provide such basic capabilities, then intervention in such a state is justified.

All accounts of humanitarian intervention have two components: on the one hand, there is a 'right to intervene' possessed by certain states, which justifies the act of intervention, and on the other hand, a condition in which the host state is in that justifies the intervention or the exercise of the right to intervene. As we saw, all accounts have different approaches to explaining the conditions under which the right to intervene can be exercised For Mills, it is the 'barbarity'; for Rawls, it is the disorderliness of society, and for Nussbaum, it is the lack of capacity to build certain capabilities in citizens that serve as conditions where the right to intervene can be exercised. We can debate the tenability of the various proposed constraints as legitimate conditions, but one thing that isn't stressed enough, and which is not very easy to see, is how the right to intervene itself is obtained. Of course, the immediate grounds can be found in various statutes and laws formulated by an international body, such as the UN. But what exactly grounds those laws? Do we ultimately reach the realm of morality in pursuit of legitimate grounds? These are interesting questions about grounding and anchoring the right to intervene. However, we do not need to go into the details of various grounding schemes of these two components, as there are a few criticisms of humanitarian intervention that we should now turn to.

Responsibility to Protect

After the Second World War, new laws regarding state sovereignty were put in place, where respect for sovereignty became the focus. But by the end of the century, a few cases of mass human rights violations emerged, notably in Rwanda and Srebrenica. The international community failed to do anything about them in time, resulting in grave genocides. It was at this juncture that Kofi Annan famously asked in his 2000 speech, “If humanitarian intervention is, indeed, an unacceptable assault on sovereignty, how should we respond to Rwanda, to

Srebrenica – to gross and systematic violations of human rights that affect every precept of our common humanity?"

In response to Annan's question, the Canadian government established the International Commission on Intervention and State Sovereignty (ICISS) and proposed the principle of the "responsibility to protect" as an alternative to humanitarian intervention. It was envisioned to be applicable in a wide variety of calamities ranging from environmental problems to atrocity crimes. The motivation behind shifting the narrative from the right to intervene to responsibility is quite worth delving into. Firstly, just having the right to intervene does not prompt a state to intervene whenever it is needed. States have the right to stay neutral too, as much as they have the right to intervene. Whereas ascribing responsibility doesn't permit capable states to stay neutral. Responsibility ascription morally induces them to take necessary actions on time.

Secondly, the right to intervene has a statist connotation by proposing that states have rights when the focus should be on the protection of the vulnerable. Thus, the responsibility to protect is an extension of the right to intervene but has the capacity to prompt capable states too. It also shifts the focus from states to the victims. Although originally conceptualized as having a broad scope of applicability, the Secretary-General of ICSS in a 2005 report narrowed the scope of Responsibility to Protect (R2P) to just four types of crimes, such as genocide, war crimes, ethnic cleansing, and crimes against humanity, collectively referred to as "atrocity crimes". Their 2009 report stated that R2P is only applicable to issues outside these areas, such as HIV/AIDS, climate change, etc., if member states have a consensus. Any attempt to extend the principle to other kinds of calamities outside atrocity crimes is thought to undermine the 2005 agreement.

Like humanitarian intervention, the doctrine of R2P has two components. Responsibility on the one hand and various calamities on the other hand. At a superficial level and for normative practices, calamities are defined in a certain way, and a definite number of conditions are considered a calamity. Out of the four crimes, three are defined in international law, and they are codified in the Rome Statute of the International Criminal Court. The third one, ethnic cleansing, is not defined as a crime in this statute; however, it has been defined by the UN as "a purposeful policy designed by one ethnic or religious group to remove by violent and terror-inspiring methods". Similarly, responsibility can be understood to be rooted in the existing laws related to sovereignty. The statutes formulated by various international agencies serve as grounding conditions for both the responsibility to protect and the atrocity crimes. The frame, which refers to the domain or space of these grounding conditions, is established through collective acceptance from member states. While this collective attitude or disposition implies that states show a certain degree of readiness to prevent certain types of crimes, it is worth considering whether this is the ultimate basis for responsibility.

Duty to Intervene

James Pattison (Pattison, 2013) proposes the model of duty to intervene as a further extension of the doctrine of responsibility to protect. He outlines various arguments for the duty to intervene, followed by objections to it, and then situates the notion of the duty to intervene in the responsibility to protect framework. After pointing out the limitations of R2P, he concludes that the duty to intervene is a more encompassing principle.

Although R2P has been politically convenient (Hehir, 2012) (Luck, 2011)to tackle and prevent various atrocity crimes, it has a philosophical problem that ultimately concerns a pragmatic issue (Luck, 2011). The first problem is that the conditions under which the responsibility to

protect is invoked and practiced are unjustifiably narrow, at least philosophically. While the narrowness of the conditions might make the principle itself politically expedient, it is philosophically indefensible. There seems to be no philosophical justification for why the principle of R2P is only applicable to atrocity crimes and not to other calamities. While it focuses on political and civic rights, it excludes other important rights such as socio-economic rights. Pattison argues that, in contrast to R2P, which is concerned with protecting a limited number of rights, the duty to intervene consists of the moral duty to intervene to protect all basic human rights. Secondly, Pattison argues that R2P is statist in essence, limiting the duty to only states and assuming that only states can provide solutions to various kinds of human rights violations. He proposes that other agencies and international institutions such as the UN can play an important role alongside states. But R2P unduly assumes states to be the foundation of the principle itself. Thirdly, R2P lacks a moral element that he deems crucial for an account for responsibility.

There are various arguments in support of the duty to intervene. The first one is called the self-interest argument, which says that the duty to intervene can be in the interest of the intervener state. The interest could be in terms of potential material gains from the intervention or more ideational interests such as the values and principles the state cares about and strives to uphold. The second argument is called the positive duty argument, which claims that the duty to intervene follows from the promise that we owe some duty to those beyond borders in virtue of our shared humanity and shared principles of universal human rights.

There is a slightly different version of this argument which Pattison characterizes as relational-cosmopolitan, which says we owe a positive duty to those outside our borders under the

global institutions that set up principles of justice (Pattison, 2013). The third argument is called the negative duty argument, which states that we owe a duty to intervene to redress the previous violations of the negative duty not to harm beyond our borders. The problem of historical responsibility is an important and heavily contested topic in the debate on humanitarian intervention too. A state could be historically involved in another state's affairs, and their involvement has led up to the present crises that require intervention. Cecile Fabris (Fabre 2012) has argued that the wrongdoer state might not be in the best position to intervene for some reasons. Their intervention might prove to be ineffective because the population of the host state might resist their intervention. But they must nevertheless shoulder some burden of intervention, such as the material and financial costs of intervention. The actual prosecution of the intervention should be done by the state that has the best chance of carrying out a successful intervention. Some (e.g., (Hjorthen and Duus-Otterström 2016) are sceptical of this move because they worry that the primary bearer of responsibility escapes the worst costs of intervention, and also argue that this doesn't serve retributive justice. It is further argued that if the wrongdoer state's intervention might be ineffective because of the potential resistance they might face, how can an intervention actually carried out by any other body but financed by the wrongdoer state not be resisted?

All the doctrines discussed have a similarity, although they differ in their range of scope and the grounds of the principles. The similarity is that the obligation to act is future-directed. Agents have the obligation to an action (intervention) that is yet to be actualized, in contrast to causal responsibility where an agent is held responsible for having carried out an action in the past. One of the differences is that each of the doctrines has a wider scope of applicability

than its precursors. Pattison's ultimate defence of his duty to intervene lies in universal human rights, the protection of which is the moral duty of every capable agent.

Broadly, these doctrines posit two components about the ascription of responsibility: the undesirable state of affairs on the one hand and the capacities of the agents on the other hand. The states of affairs generate obligation from the agents, and the obligation essentially is ascribed with respect to these states of affairs. I call these states of affairs "generators." The capacities possessed by agents are what ground the ascription of responsibility to those capable agents. Any practice of responsibility ascription assumes certain states of affairs as generators a priori. The construction of grounds is usually about setting up a way of finding certain descriptive features of the world on the basis of which responsibility ascription to particular agents is justified.

While Pattison's notion is an improved extension of the previously proposed principles of military intervention and responsibility to protect, it does not hold itself up to the limits of applicability that our contemporary intuitions regarding demands for action require. In the contemporary world, the rights of animals are an active topic of debate, as well as a concern for policymakers. So are the issues related to ecology and the environment, such as climate change. Although Pattison doesn't suggest that the duty to intervene is applicable to animals and environmental issues, that is, all the calamities in the world should be understood in the framework of human rights, his proposal does not keep the possibility of an alternative approach open either. Even if all the calamities are included in the human rights framework, it will be clearly problematic. The doctrine will be acutely anthropocentric in its formulation. Just like humanitarian intervention is statist in its approach compared to the responsibility to protect, the duty to intervene (along with its precursors) is anthropocentric. It undermines

the rights of animals and non-human beings. Another potential question that can be targeted at all these doctrines is whether it is correct to use violations of certain rights (such as human rights, animal rights, etc.) as a legitimate condition for the exercise of the duty to intervene. That is to say, is it even correct to talk about duty within the framework of 'rights'? Following the proposed doctrines, we might be tempted to posit a certain right borne by the environment, the protection of which is our responsibility. But is that reasonable, or at least, an effective way to conceptualize the grounds of the kind of duty we are interested in?

Therefore, intervention responsibility as I define it is proposed as an extension of the doctrines mentioned above. It has a much broader scope of applicability because it covers issues like ecology and climate change in addition. Moreover, unlike most of the doctrines discussed which assume statism as their ontology (which is criticized by people like Pattison for being too narrow and potentially less effective), intervention responsibility doesn't take a simplistic model of governmental system, with only states as its entities, as its ontological framework. While intervention responsibility is defined to be applicable to a wide range of problems, the discussions around this notion in this thesis will be confined to the context of climate change. Towards the end of the previous chapter, I had hinted at some shortcomings of the existing models of forward-looking responsibility. In particular, I pointed out that the capacity to intervene, on which the entire notion of forward-looking responsibility often hinges, is often left unelucidated. Furthermore, I argued that while a forward-looking approach does not entirely avoid causal notions, it hardly provides a comprehensive account of them. Some generic causal notions that are almost always used are mitigation, influence, intervention, remedy, etc. I now turn to the main task of this thesis, which is to give a rigorous account of the capacity to intervene and the process of intervention. To this end, I will draw

upon the rich resources provided by the causal modelling and interventionist framework of causation.

Chapter 4
Formalization of Intervention

While the philosophical literature on causation, as discussed in the literature review, mainly concerns the semantics and metaphysics of causation, another approach (or a set of approaches) has gained significant traction in contemporary debates on causation. This tradition of research emerged in the late 20th century, largely in lockstep with the development of computers, and takes influence from varied disciplines, such as econometrics, computer science, epidemiology and other related empirical fields of inquiry.

The dominant methodology throughout the 20th century was statistical inference. In essence, all such studies are observational, which means that inferences are drawn from mere observation. The primary aim of such studies is to measure the probabilistic dependence between different variables of interest. This probabilistic dependence between two variables is called correlation, and the variables associated in this way are said to be correlated. Correlation is obtained if there is a co-occurrence of two events (or variables). In other words, if the probability of a variable taking a particular value given that another variable takes a specific value is higher. The higher the probabilistic dependence, the stronger the correlation.

For a long time, there was almost a general consensus among scientists and social scientists that the aim of their research must be to find correlations between different variables. The correlation was easy to compute mathematically from data, and it was very useful for the purpose of prediction. If you see one event happening, then you can predict that the other event will happen too, given a correlation is established. So, it seemed very plausible that this

relationship is all there is to worry about for the practitioners of these fields. While it is based on seeing/observing the world and inferring patterns or generalizations from them, it misses one crucial aspect of reasoning that humans have developed over the ages, which is what makes humans draw fairly reliable inferences. And that is the concept of action. Observational studies rely on the gathering and analysis of data and place no importance on action. Causality involves action in the process of drawing inferences, and that's what makes it different from correlation.

In correlational studies, there is no way to actually see if the patterns and generalizations persist if a human manipulates the data. This methodological deadlock exists due to the disregard for causation in such studies. While such studies can narrow down the data and rule out the space of possible conclusions, it is extremely hard and mostly impossible to tell us if there is a causal relation between two variables. Suppose we see a correlation between variables X and Y. From this information, we can certainly make predictions; every time we see X changing its value, we can expect Y to change its value. But does that mean that X and Y are causally related? To put it differently, does the dependence between X and Y persist if the value of X were to be changed? The answer is not straightforward.

Another crucial difference between correlation and causation is that causation entails counterfactual dependence, whereas correlation does not. Event A is said to cause event B if B always follows A, and if A had not occurred, B wouldn't occur either. The first part of the definition merely describes the correlation that is observed. The second part, however, puts a critical condition on the correlation. The correlation observed cannot just be accidental, but it must be instantiated by an underlying mechanism. Event A produces or brings about event

B through a mechanism, such that if the cause had not occurred in a possible world, a mechanism would not have been generated, and the effect would also not follow.

The need to find causation was partly driven by the desire to manipulate systems and provide explanations for various phenomena, which are not provided by correlation. The late 20th century saw many contributions from distinct fields, all of which had differing motivations behind the need to understand and capture causality. The impediment to various efforts was the lack of a formal language to capture causality in mathematical terms, unlike correlation, which had been developed very richly throughout the 20th century. However, an influential book written by the computer scientist and philosopher, Judea Pearl, that came out in 1999, changed the course of history for causation. Pearl had realized that for computers to reach human-level intelligence, causality was the first and basic element of the human mind that must be incorporated in it. Causality is the language of learning for humans. It is the language that infants use too, making it very innate to human minds. And it is a language that sets humans apart from all other living beings. Judea Pearl came up with a formalism for causality and set up a method to infer causal relations from data known as causal model. The topic of the next section covers the basic features of the causal model framework.

Causal Modelling

Causal modelling is an interdisciplinary field of research that developed with contributions from various fields, such as econometrics, epidemiology, computer science, and philosophy. A causal model is a mathematical model that represents causal relations in a system. The collection of all causal relations in a system is called the causal structure of the system. It helps us make predictions about our actions; what would happen if we intervene on particular

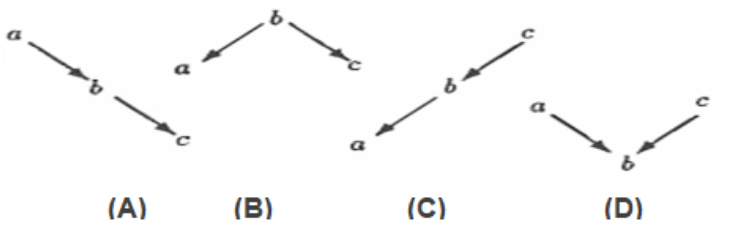
causal relata. It can also help us infer the probabilistic dependence of various variables on each other. Conversely, it can help us find the correct causal relation from probabilistic dependence or even the whole causal structure from the probability distribution (Sloman, 2009). Therefore, in contrast to correlation, which tells us the association between variables, it gives insight into the mechanism that produces correlation. With the help of the mechanism, we can intervene and manipulate to change the outcome or the whole structure too. Not only does it give us the result of a feasible intervention, but it also tells us what the results of a hypothetical intervention can be (Woodward 2003).

Causal models have three components (Sloman, 2009). The first component is the system that is assumed to work causally. The second component is the probabilistic distribution of the system, which tells us how the variables are conditionally dependent on each other. That is to say, how likely they are to occur together in combinations. The third component of a causal model is the causal graph, consisting of directed arrows and nodes. The nodes represent variables and the directed arrows represent causal relations, which tell us which particular variable is causing which other variables. In this framework, both the probability distribution and the causal graph represent the world. And the causal graphs are also representations of the probability distribution. The causal graph (which is the representation of the structure) restricts the probability distribution. In that, it constrains how a probability distribution can look like. If there is no arrow between two variables, then there is no probabilistic dependence, thereby simplifying the set of data. In this way, the causal graph produces the probabilistic nature of the world that we observe. If the graph is missing a particular arrow, or a cause is unknown (and unrepresented in the model), we find probabilistic dependence between them. However, it does not rule out the deterministic causal relations in the world.

Events or variables can be deterministically related in a causal structure too when the dependence isn't probabilistic (Sloman, 2009).

How causal structure produces the probabilistic distribution

If we find a correlation between two variables or events, what causal relation can we infer from this? From this information, one can construct multiple equivalent causal graphs. Suppose ***a*** and ***c*** are correlated. Then either one could be the cause of the other, or the causal structure might look like one of these graphs in the picture below. Any of these graphs is perfectly compatible with the information that ***a*** and ***c*** are correlated. If it isn't a straightforward causal relationship between ***a*** and ***c*** with just one arrow pointing from one to the other, we are missing a variable in the structure.



The causal modelling framework provides tools to figure out which graph exactly represents the data. To find the cause of a variable, we condition on every variable except that particular variable which we are trying to check if it's the cause by freezing them at a particular value (Pearl, 2009). Then we intervene on that variable by changing its value. If we see an associated change in the other variable, then it is the cause. In this way, this framework provides a way to represent intervention and helps us predict the result of our action, unlike correlation which just represents what we see. To represent intervention, Pearl (Pearl 2000)introduced a mathematical operator called the do-operator. To represent the intervention on a variable

X by fixing it at a value x, we apply *do(X=x)*, which is then used to calculate the interventional probabilities. Graphically, intervention is represented by cutting off all arrows that come into the variable that we are intervening on. Pearl calls this graphical surgery.

In graphs *A* and *C*, we have chain graphs, where ***a*** and ***c*** are correlated, and ***b*** is the mediator. If you condition on ***b***, ***a*** and ***c*** become conditionally independent. This is called the Markov condition or screening off. Here, ***b*** screens off ***a*** from ***c***. If you have information about ***b***, then the correlation between ***a*** and ***c*** disappears. For example, let's say ***a*** is a lightning strike, ***b*** is ***a*** fire, and ***c*** represents heat. One can find a correlation between lightning and fire, but if fire is conditioned on and we intervene on lightning, there will be no change in ***c***, which implies that ***c*** is independent of ***a*** causally. Graph *B* represents a causal fork, where the correlation between ***a*** and ***c*** appears due to a third variable which is the common cause of ***a*** and ***c***. Suppose ***b*** represents fire, ***a*** is the heat generated, and ***c*** is the damage that occurred. If we condition on fire, then the correlation disappears between heat and damage. They are said to be independent, conditional on ***b***. That is to say, if we know that there is or is not a fire, then the information that there is damage or no damage does not tell us anything about whether there is or isn't heat. The third kind of a basic graph is the collider or inverted causal fork. Figure *D* represents an inverted fork. Here, ***a*** and ***c*** are initially independent, but conditioning on ***b*** makes them dependent. These graphs are called Markov equivalents because they all can satisfy the same probabilistic dependence relations. These tell us that correlation inferred from a set of data can give us confusing and problematic results. It takes intervention to know exactly which is the correct graph.

Interventionism

The motivation behind the development of causal modelling was mainly to find a technique to see the results of manipulation and intervention. This is rooted in the role of experimentation. The research in causal modelling had an enormous influence on the philosophy of causation, although the researchers in these fields were mainly concerned with developing a quantitative method of causal inference. James Woodward's work (Woodward 2003) first brought these developments into the philosophical literature and shed light on the philosophical understanding of causation under this framework. The theory of causation he developed is termed as interventionism or more broadly, the manipulability theory of causation. Although it does not illuminate the metaphysics of causation, it mainly concerns the epistemology and methodology of causation. It is not supposed to be a reductive theory of causation, that is, it does not attempt to define causation in terms of non-causal primitive notions. For instance, the notion of intervention itself is a causal one. But it can nevertheless be illuminating (Woodward 2003) by elucidating various sorts of causal claims such as total cause, direct cause, actual cause, and contributory cause. It is interpretive and semantic, in contrast to the works on causal inference, which are quantitative. It explains in considerable depth the meaning of the intervention variable and the conditions that have to be met for a process or event to be an intervention. Prior to the work of Woodward, Price and Menzies had developed a manipulability theory of causation, where causation was defined in terms of human agency and free action. The employment of human agency and free action was an attempt to give a reduction account of causation by appealing to a primitive notion of free action and agency. Human agency was an essential and primitive idea, which was experienced by the agent and needn't be explained further. In this theory, causation is defined as: "An

event A is a cause of a distinct event B, just in case bringing about the occurrence of A would be an effective means by which a free agent could bring about the occurrence of B" (Menzies and Price 1993:187–203).

Reference to free action is supposed to imply that there was an intention from the experimenter in the process of manipulation. Since human agency is something that we have direct experience of and which plays a fundamental role in our experience and perception, this account was proposed as a non-circular reductive one. However, human intervention is not possible in every situation. There are many experiments in which human intervention is either physically impossible or problematic for ethical reasons. Since human agency plays a fundamental role in this definition of causation, it has been criticized as being too anthropocentric. In contrast to this, Woodward (Woodward, 2003), following Sprites and others (Spirtes et el. 2000), characterizes causation in terms of a hypothetical manipulation or intervention. In this approach, inferences are drawn from a hypothetical experiment where the intervention is carried out by a hypothetical process in a possible world. Although intervention can be carried out by humans too, it's not necessary in principle. Experiments involving human intervention are only a subset of all the experiments that can tell us about causal relations. In this way, by using an aspect of Lewis's counterfactual account of causation, it evades the criticism of anthropocentrism. Similar to the idea of intervention as a graphical surgery proposed by Pearl (Pearl, 2000), Woodward (Woodward, 2003) compares intervention to pressing a switch. The idea is that the switch blocks all the causal influences on the intervened variable. The intervention variable *I*, which is an exogenous variable that intervenes on a variable in the model, can assume different values, and not all the values of *I* can act as a switch. There are only a particular set of values assumed by *I* that can disconnect all the causal influences or arrows coming from a set of variables into a variable X on which

the intervention is to be carried out. Suppose there is a set of variables V that are causally relevant to X, and there are some values of *I* that block all the causal influences on X, then *I* is said to be an intervention variable relative to V. An example from Woodward (2003) elaborates this notion of *I* as a switch.

> Consider a stereo receiver that has dials for volume, treble, bass and a power button. The setting of each is causally relevant to the frequency and amplitude of the sound waves that are emitted from a speaker. However, the settings make a difference to the sound emitted only when the power button is in the "on" position. When the power button is set to "off," there are no possible changes in the position of the other dials that will change the output of the speaker. Under this condition, it is as though the causal connection between the position of the other dials and the output of the speaker is broken. The position of the power button is thus a switch for the output with respect to the other dials. (Woodward 2003:96)

When the switch (I) is on, it cuts all the arrows coming into X. And it is only certain values of (I) that have the capacity to cut the arrows.

Woodward characterizes a very general notion of intervention and develops a set of conditions that the intervention variable has to meet:

"I" is an intervention variable for "X" with respect to "Y" if and only if "I" meets the following conditions:

1. "I" causes "X".

2. "I" acts as a switch for all the other variables that cause "X". That is, certain values of "I" are such that when "I" attains those values, "X" ceases to depend on the values of other variables that cause "X" and instead depends only on the value taken by "*I*".

Given the notion of an intervention variable, an intervention may be defined as follows:

(IN) I's assuming some value "I = z" is an intervention on "X" with respect to "Y" if and only if "I" is an intervention variable for "X" with respect to "Y" and "I = z" is an actual cause of the value taken by "X" (Woodward 2003:98)

Although intervention has to be thought of as a hypothetical process for generality, for our purposes, it will suffice to consider the subset of this notion where the intervention can actually be carried out by humans. Since we are not defining causal claims in terms of agency, we do not have to worry about anthropocentrism. We are only concerned with those causal relations which are manipulable by agents.

If the intervention is carried out by a human agent, then the intervention variable must be thought of as an agent. The range of values taken by the intervention variable corresponds to the range of capacity of the agent.

Intervention Agency

Intervention understood in the narrower sense is a representation of action by an agent. The notions of action and agent play a crucial role in the elucidation of intervention, thus requiring discussion. These two notions are very closely related, and they have been deeply discussed in the philosophy of action and related areas. The question that should be asked here is: what exactly counts as an agent, and what counts as action? What makes an action different from a mere movement? Agency, broadly construed, can be thought of as possessed by everything

that is causally efficacious. In other words, everything that plays a role in a causal structure, in some sense, possesses some sort of agency. And the sophistication of this agency varies according to the organizational complexity of the entity. But this characterization of agency seems too general, unilluminating, and unfruitful for normative purposes. A narrower characterization of agency, which traces its origin at least to Aristotle, is more influential. It defines agency in terms of the intentionality of intentional action or action generated by intentional action. The literature on the philosophy of action provides a standard conception and theory of agency. Both the conception and the theory have criticisms; however, the conception seems to be broadly accepted. The conception reduces agency to intentionality, and for a long time, the term intentionality had largely replaced agency in the literature. In debates on moral responsibility, this kind of agency, let's call it intentional agency, was a useful tool in assigning retributive responsibility. An agent who committed an offense intentionally was to be held responsible. However, in forward-looking responsibility, this kind of agency seems insufficient, although required, for the assignment of responsibility. It is necessary because the ability to intend to pursue a goal by following a strategy is required to achieve the goal successfully. But before intending to follow a strategy and pursue a goal, one must be capable of setting up a goal and a strategy to achieve it. The standard conception considers agency to be the exercise of intentional action, and intentional action is that which is caused by certain mental states. Thus, there is a way to rationalize intentional action from a first-person perspective by appealing to certain mental sets, such as desires and beliefs. But what stops us from defining agency in terms of those mental states as opposed to intentionality, which is less fundamental than mental states? My speculation is that intentionality gives a functional solution to the problem of the assignment of retributive responsibility.

What then can ground forward-long responsibility like intervention responsibility? Clearly, we can't just assign intervention responsibility to any agent who is capable of forming an intention. I had briefly intimated about the features that set humans apart from other non-human beings. Humans are said to have acquired a powerful cognitive capability some thousands of years ago. This event is called by some historians the cognitive revolution and resulted in the rapid progress of humankind. This newly acquired cognitive capacity consisted of the ability to imagine non-existent things, make assumptions about them, and derive conclusions. This is what we call counterfactual reasoning. It was an enormous leap in the method of reasoning and resulted in significant changes in humans. Counterfactual reasoning, which entails causality, replaced the previous method of reasoning from observation, which we shared with many other living beings. Humans became capable of working as teams, making strategies, and assigning roles to carry out activities. This capability involves the ability to represent causal structures. For example, a group of people planning to hunt a prey could imagine causal models of the scenarios and make inferences by conditioning on various variables that might influence the outcomes. Over time, humans started imagining more and more sophisticated causal models that helped them achieve impressive goals. So, to answer the question of what the ability to make strategy consists of; it is the ability to represent the causal structure. With the help of causal structure, we are able to represent intervention. The ability to represent causal intervention is thus the intervention agency (Sloman, 2009). The causal models can range from rudimentary models that we make in our daily lives to the more complex ones that are constructed by scientists. An intervention agent is a kind of agent that can represent causal models and interventions. The representation tells the agent what the mechanisms that drive various systems are, how things behave with each other, and more

importantly, how things would behave if we set different initial conditions and values of the variables.

Agency Inside a Causal Model

The causal modelling framework treats the intervention variable as an exogenous process that causally interacts with the variables inside the model. However, for normative practices like responsibility ascription, an objective way to represent agents and assess their responsibilities is required. To assess the responsibility of various intervention agents in the world, we need a way to calculate certain descriptive features of the world as well as the agents that are relevant for the ascription of responsibility. To estimate the responsibility by assessing features that are both intrinsic and extrinsic to the agents, and compare the responsibilities of various agents in the world, we need a way to represent both the agents and physical variables in the same framework. A framework that provides us with insights and tools to calculate these relevant features. The relevant features are essentially causal efficacies of different agents and physical variables. An off-the-shelf framework to represent these is a causal model.

A causal model represents a system of cause-and-effect relationships between different variables, where each variable can be influenced by other variables in the system. An agent can be seen as a variable in this system, and its actions or decisions can be seen as inputs that influence other variables and the system as a whole. There are two kinds of variables in such a model: the first is the agential variables that represent agents, and the second is the non-agential variables that represent physical variables. These variables are all causally efficacious and represent the ability to make changes in the model. Graphically, the agents reside in the

causal graph, occupying definite nodes, and arrows pointing into and out of these agential variables. The agential variables differ from other non-agential variables by virtue of their possession of intervention agency. An example to illustrate the importance of the inclusion of agents in a causal model is the following: let's say we have a causal model that represents a traffic system, where different variables such as traffic lights, road conditions, and driver's behaviour all impact the flow of traffic. An agent in this system could be a driver who makes decisions about when to accelerate, brake, or change lanes. In such a case, the flow of traffic is not just affected by variables such as the condition of the road, traffic lights, etc., but also the actions and decisions of each individual driver. To understand how an agent's actions or decisions impact a causal model, it's important to consider various causal pathways within the system. For example, an agent's decision to accelerate may lead to increased traffic flow initially, but if too many drivers make the same decision, it could lead to congestion and slower traffic later on. Therefore, the inclusion of the agents inside is an important part of the analysis to have a comprehensive understanding of the model.

It is important to distinguish two kinds of representation here. One is the ability to represent a causal model from the first-person perspective where the agent itself is exogenous to the model. This representation is possessed by the agent who actually intervenes on a non-agential variable, such as the emission of GHG. The ability to represent from a first-person perspective essentially is what makes an entity an agent. And it is one of the crucial features that explains why an entity is responsible. The second kind of representation is useful for the purpose of responsibility ascription, which is itself a kind of intervention (Hitchcock and Knobe, 2009). In this model, interventions on other agential variables are represented. To ascribe responsibility to an agent that resides in a model amounts to inducing that agent to

causally interact (intervene) in a particular way so that the outcome is desirable. In this second kind of representation, all the intervention agents are included in the model. From this, the degrees of responsibility can be evaluated too. The first kind of model is plainly descriptive, and the second kind is an extended version of the first - because it includes all the intervention agents too - while also having a normative feature.

Chapter 5
Defining Intervention Responsibility

The notion of intervention responsibility can be defined, to a first approximation, as a moral status that is ascribable to an entity possessing certain capacities and powers with respect to a state of affairs that is undesirable. Whenever those undesirable states of affairs obtain, the agent is responsible for the exercise of their power and capacities to make it less undesirable. The nature of this capacity and power is partly discussed in the previous chapter, and the nature of undesirable states is discussed in the chapter before that. However, notions like the capacity of the agent and the undesirability of outcomes need to be discussed further and elucidated on the way in which they provide the grounding conditions for intervention responsibility. But before that, intervention responsibility will be defined in more concrete and formal terms, using the framework of causal modelling and interventionism. Additionally, the nature of states of affairs, which I called generators, will be discussed again towards the end of this chapter.

The interventionist framework (Pearl, 2000) (Woodward, 2003)has been extended to understand the attribution of causal responsibility (Halpern and Hitchcock 2015)(Hitchcock & Knobe, 2009). Since the work of Halpern and Hitchcock in moral psychology and experimental philosophy, the literature in this tradition has grown. Experimental studies on actual causation have revealed that assessments of actual causation are influenced by various factors such as the typicality, normality, and defaults. Studies conducted by (Samland and Waldmann, 2016) and (Icard et al. 2017) have all demonstrated this variability. Another study has shown that people's judgments of actual causation is influenced by their understanding

of norms (Hitchcock and Knobe, 2009). Some studies have concluded that the degree of causal redundancy also affects judgments of causal strength and responsibility. In situations where multiple agents contribute to the same outcome, such as when two marksmen shoot a person, the responsibility of each agent decreases as the outcome becomes more overdetermined, as seen in the study conducted by (Lagnado, Gerstenberg, and Zultan 2013). Same study (Lagnado et al. 2013) also shows that judgements of intuitive responsibility are function of pivotality and criticality of agents with respect to an outcome. Another study concludes that judgments of causal responsibility are influenced by the robustness of the causal chain linking the agent's action and the outcome (Grinfeld et al. 2020). The robustness tells us how stable the causal relations are with respect to the background conditions. The more stable the relationship is as we change the background conditions, the more robust the causal relationship is.

While these studies focus on the intuitive judgments of responsibility ascriptions and how they are affected by various properties of the causal structures, (and how judgements about the causal structure, such as actual causation, are in turn affected by various normative factors), they don't really posit new or modify traditional notions of responsibility. They take a particular notion of responsibility, say for example, causal responsibility, and attempt to find the properties of the world, or at least the properties of the world as perceived by humans, that affect our judgments of causal responsibility. Another feature of these studies is that they are all essentially backward-looking. Causal responsibility as understood traditionally is a backward-looking notion where the causal relation is already obtained in the world. That is, the relata should have instantiated, and the relation established to do the analysis and ascribe responsibility.

Therefore, while these studies can illuminate us on the descriptive features, or at least as perceived by us, that shape our intuitive judgments of responsibility, they can't provide us with a way to give an account of a forward-looking notion like intervention responsibility.

The interventionist framework (Pearl 2000; Woodward 2003) has not been extended to understand a primarily forward-looking responsibility so far. But it seems plausible that we can potentially extend the interventionist framework to give an account of intervention responsibility that I'm proposing in this thesis. Before I try to explicate the notion of intervention responsibility using the interventionist framework, a couple of additional modifications and extensions of the original framework are in order. The modifications are not meant to posit any metaphysical propositions about causation. Rather, they pertain to either constraining certain notions in the original interventionist framework or extending the framework to the realm of sociality and morality.

The first modification pertains to the notion of intervention itself. The interventionist framework characterizes intervention as an exogenous process that changes the value of a variable, which further triggers a change in another variable. The change here could be either positive or negative. In other words, the intervention could either raise or lower the probability of the effect by changing the probability of the cause. The variable or state of affairs, which we expect to change after our intervention, possesses no intrinsic features such as desirable or undesirable. Since our concern is normative, we will have to ascribe such statuses of desirability in our model, thus extending the original framework. Additionally, the kind of change that we must expect is not any arbitrary change. If the state of affairs is undesirable, then we expect the intervention to lower its probability of happening. If it is desirable, then we expect our intervention to increase its frequency. In short, if a state of

affairs is undesirable, then intervention must be negative. If a state of affairs is desirable, then the intervention must be positive. Although, in our discussion on responsibility, we are mainly concerned with only those states of affairs which are undesirable. We are concerned with climate change in this thesis, and climate change is taken as an undesirable state of affairs.

The second modification pertains to the idea of the intervention variable, referred to as the intervention agent in our account. In the traditional accounts of causal responsibility, an agent who is the cause of a certain state of affairs (usually a harm) is attributed a responsibility. The cause here, which is the action of the agent, must have already occurred. What, then, must we refer to the intervention agent as? Clearly, none of the notions of cause, such as actual cause, contributory cause, etc., can be ascribed to the agent. The agent here has the potential to be the cause, yet the causal link between the agent and the variable has not been actualized. I call the intervention agent the potential cause. This line of characterizing the intervention agent can be potentially explored further. For now, I will focus on the explication of intervention responsibility.

The notion of intervention responsibility can be understood as follows: Agent A is intervention responsible for a state of affairs S (or a variable taking a certain value) if there is a possible intervention on the cause X of S for the agent A, such that intervening on X would be a way to manipulate S. This is a slightly general formulation of intervention responsibility as it does not constrain the nature of the state of affairs. A narrower formulation for our purpose can be formulated as follows: Agent A is intervention responsible for an undesirable state of affairs S (or an undesirable variable taking a certain value), if there is a possible intervention on the cause X of S for the agent A, such that intervening on X would be a way to mitigate S.

Grounding conditions for Intervention-Responsibility

The grounding problem about intervention responsibility basically amounts to asking the question: What explains the ascription of intervention responsibility to a particular agent? As mentioned earlier, intentionality seems to be an important notion to ground intervention responsibility too. An agent who is incapable of forming intentions, even with the possession of relevant reasons, to carry out a certain task seems like an improper site of intervention responsibility. The intervention agent must be goal-directed, and thus intentionality is a necessary grounding condition of intervention responsibility, like any traditional notion of moral responsibility. Another ground of intervention responsibility is moral competence. As discussed earlier, intervention responsibility is ascribed to an agent with respect to a particular state of affairs, and the state of affairs that demands intervention responsibility has a certain moral status. For an intervention agent to be intervention responsible, it must be capable of perceiving the moral significance of that state of affairs. The moral status borne by a state of affairs that generally demands obligation from an agent, and thus intervention, is moral faultiness. For example, for an agent to be an intervention agent and deserving of intervention responsibility attribution for global warming, it must be capable of understanding that global warming is undesirable and morally bad. The undesirable states of affairs that involve morally faulty conditions are the generators, which I discuss in the next section.

The set of grounding conditions that I'm most concerned with in this thesis isn't intentionality or moral competence. Here, I'm going to explicate the notion of intervention capacity, which I propose as an aggregate measure of two important grounding conditions , both of which are absolutely necessary for the ascription of intervention responsibility. The first one is the notion of intervention agency that I have already discussed. It is quite clear that the

possession of intervention agency is a necessary condition for the ascription of intervention responsibility. A person lacking the ability to make sense of the world by understanding various events and mechanisms that drive them is not liable for the ascription of intervention responsibility. Intervention agency is an essential feature of the agent and something that makes agential variables and non-agential variables (like GHG emissions) different. An agent must be able to represent the causal structure through a causal model and various interventions in it, and infer the effects of hypothetical interventions. For example, a passer-by who sees a house catching fire must have the ability to represent the causal model of the scenario and infer that fire can cause damage to properties, kill people, and pollute the air. She must also be able to reason counterfactually about the scenario; if she were to intervene, then the course of events that usually follow from that state of affairs (without intervention) can change too. The agent is intervention responsible in this situation to take action in its capacity by virtue of having these abilities. The requirement of intervention agency is essentially an epistemic condition on the ascription of intervention responsibility.

The second grounding condition for intervention responsibility is the notion of intervention strength. It is a feature that, although possessed by the agent, is conditioned by the causal structure. Intervention strength can be qualitatively characterized in two ways. Firstly, intervention strength can be thought of as the range of values assumed by the intervention agent, to which a causal structure is sensitive. It is those values assumed by the agent that have the capacity to intervene on a variable by blocking all the causal influences on that variable and manipulating its effects. In other words, it is the capacity of the agent to successfully carry out an intervention. An entity is an intervention agent only relative to a set

of variables V that are causally relevant to X in a causal model. Thus, an entity is intervention responsible for X if they are an intervention agent on X relative to V.

In practical cases, the values assumed by an entity or the intervention strength could be the aggregate of the number (and their amounts) of skills, resources, etc. possessed by an agent that has the capacity to manipulate the effect of a variable. Note that although the strength is possessed by the agent, it is conditioned by the causal structure. The exact skills and resources, as well as their precise magnitudes, that are required for a successful intervention are conditioned by the causal structure itself.

An example to illustrate this characterization of intervention strength would be the following: Suppose a person sees a child drifting in the middle of a river. For this person to be intervention responsible for rescuing the child, firstly, she must possess intervention agency. Secondly, she must have the required intervention strength. The intervention strength of this person is the aggregate of her level of swimming skill, physical strength, and so on. Although the swimming skills and physical strength are possessed by the intervention agent, the required amount of this aggregate quantity is constrained by features of the causal system, such as the flow of the river, depth of the river, temperature of the water, etc.

Another way to characterize intervention strength is to appeal to the potential effectiveness of the required intervention. That is, how potentially reliable that intervention is in producing the desired effects in the system. The effectiveness can be defined in terms of determinism. It can be thought of as how deterministic a particular intervention is. For an agent to be intervention responsible to mitigate global warming, the causal relationship between the intervention agent and the net atmospheric GHG (nAGHG) must be deterministic (or at least

to a certain extent), and the intervention agent's action must reliably produce the desired changes (that is, reduction in GHG emissions).

The determinism of the causal relation between the intervention agent and GHG emissions can be further defined in more familiar terms of sufficiency. For a causal relation between the intervention agent and nAGHG to be deterministic, the intervention agent must be sufficient to produce the desired effects. And for that relationship to be absolutely deterministic, the intervention agent must be completely sufficient for the effect. However, sufficiency can come in degrees, and there should be a way to evaluate the degree of sufficiency. The precise and formal way to capture sufficiency will be left as an open question.

Both of the characterisations can then be used to evaluate the intervention-strength of an agent. Although, I have provided a qualitative way of assessing the intervention-strength, I don't dismiss the possibility of there being a more rigorous and formal way to capture the intervention strength. Together, intervention strength and intervention agency are proposed as grounds for the ascription of intervention responsibility. An entity is intervention responsible for mitigating an undesirable state of affairs (which entails moral bad in the world), if that entity possesses intervention agency and the necessary intervention-strength. Therefore, the ascription of intervention responsible requires the establishment of, as a matter of fact, the intervention-agency and intervention-strength. Once an entity is understood to possess the ability to represent causal structures or causal models (and consequently the ability to represent intervention) and the ability to actually carry out the intervention, we are justified in ascribing intervention-responsibility to that entity.

Generators of Intervention Responsibility

Another component in the ascription of intervention responsibility pertains to the nature of the state affairs with respect to which intervention responsibility is ascribed. The state of affairs must be taken to entail a moral bad, which is generally assumed prima facie. For example, climate change causes misery and harm to living beings and nature, and causing (either by action or omission) misery to living beings and nature is morally bad. Morally bad states of affairs must be mitigated (or eliminated, if possible), and it is these morally faulty states of affairs that demand intervention from those agents who are intervention responsible. The morally faulty states of affairs trigger a relation to be generated between the intervention agent and the state of affairs. As I explained in the third chapter, this component is the generator of intervention responsibility.

GHG emission causes global warming, and global warming causes various sorts of misery. The miseries caused by the emission of GHG entail morally faulty states of affairs, and these morally faulty states of affairs can be categorized into three broad kinds (Gardiner, 2011). The first is the violation of human rights. The people who are most vulnerable to climate crisis and who lack the basic means to adapt to the changing climate and other effects of climate crises can be understood as being robbed of their basic human rights to live a decent life. The import of this approach becomes more apparent when we do socio-economic analyses of the victims of the crises. Those who lack the basic means to adapt are generally the ones who hardly contribute to the crises in the first place. The literature on climate justice sees climate crises as rich people appropriating a disproportionate share of global public good at the expense of poor people by causing disproportionate harm to the poor (Jamieson, 2014).

Under the broader framework of global justice, it can be seen through multiple lenses. I propose that a comprehensive and intersectional ontological framework including states, corporations, jurisdictions, organizations, and collectives (such as race, caste, and other strongly glued social groups) must be assumed, which is to be formulated as rich countries, states, jurisdictions, races, and castes appropriating unfair amounts of the common good while disproportionately harming poorer and marginalized ones.

Another kind of justice that is invoked in the literature is intergenerational justice. Climate change is caused by us and the generations that lived before us, but the cost will be borne by the generations yet to come (Gardiner, 2011). The third, and often underemphasized, issue that climate crises involve is the idea of ecological and environmental justice. Climate change can't be just seen as posing threats to humans alone. Ecological justice is a broader framework that emphasizes the primacy of all living beings, including the environment, and the need to take the entire ecology into consideration while deliberating about any sort of climate action. Consequently, the all-encompassing framework of ecological justice also escapes the problem of anthropocentrism. In this framework, climate justice isn't just seen as a human-rights issue but an issue of the moral status of non-human beings too.

However, a problem that I had first discussed, although briefly, towards the end of the third chapter about the ways to look at environmental justice still needs some elaboration. In particular, the question I raised about the exact generator of intervention responsibility within the framework of justice. I have attempted to provide some answers to this question above in this section: Climate crises are an issue of global justice, intergenerational justice, and ecological justice.

While the first two kinds of justice are fairly straightforward once we accept the framework of human rights, humans in the world have certain basic rights, such as the right to live a decent life with access to basic necessities, and so do the humans who are yet to come into the world in the future. The protection of those rights is the responsibility of agents who possess certain capacities and powers. More precisely, it is the responsibility of those who possess intervention agency and intervention strength to intervene whenever the rights of victims (who are currently alive or those who are yet to be born) are violated.

The idea of ecological justice is relatively less intuitive because it posits the existence of certain intrinsic values and entitlements to all living beings. But it nevertheless has great functional appeal. In virtue of these intrinsic values, all living beings are understood to have moral standings. Any threat to their lives, well-being, or habitat is to be seen as a violation of these entitlements. These moral entitlements or intrinsic values, whatever their right characterizations are, attributed to living beings, generate the obligation in humans. Whenever an instance of the violation of their moral entitlements obtains, humans are obligated to intervene by rectifying the harm caused to them, thereby protecting their moral entitlements. Notwithstanding its less intuitive nature, the framework of ecological justice can play important functional roles in the overall success of climate action because it is far more comprehensive and takes every living being into consideration.

Finally, to define intervention responsibility more precisely, agent A is intervention-responsible for mitigating an undesirable state of affairs S (or the variable that produces S taking certain values) if:

1. A has strong enough intervention-agency.

2. A has enough intervention strength to intervene on the variable by cutting off all causal influences on S and changing its value so that there is an associated positive change in S.

3. S is a proper generator of intervention responsibility.

To summarize, the ascription of intervention responsibility involves the establishment of two sets of conditions. The first condition is that there must be a generator (there can be multiple generators too). These are notions that generate intervention responsibility in agents. It has to be assumed before actually ascribing intervention responsibility to an agent. The second condition involves setting up the grounds on the basis on which intervention responsibility is ascribed to particular agents. These are features of agents and the causal structure that explain why certain agents are responsible and not others.

Some further remarks on intervention responsibility

In the previous section, I discussed various generators of intervention responsibility in the context of climate change. I essentially argued that the act of intervention in the appropriate way fulfils the demands generated by those generators. Some of these generators were global justice, intergenerational justice, and ecological justice. Intervention responsibility is a forward-looking notion of responsibility that is either aimed at mitigating present harm or preventing potential harm in the future by taking various preventive measures. And the responsibility falls on those agents who have certain powers and capacities, irrespective of whether they were complicit or not in bringing about those harms. Intervention responsibility can help fulfil the obligations of global justice because it involves recognizing that we have a

shared responsibility to ensure that basic human rights are respected and that all people have access to the resources necessary for a decent life. This can involve intervening in situations where these rights are being violated. Intervention responsibility fulfils the obligations generated by intergenerational justice by recognizing that our actions today have consequences for future generations. By intervening to address environmental degradation, resource depletion, and other issues that threaten the well-being of future generations, we can help ensure that they are able to inherit a liveable planet and a sustainable society. And finally, intervention responsibility fulfils the obligations of ecological justice by recognizing that humans have a responsibility to protect and preserve the natural environment. By intervening to address environmental degradation, we can help protect the rights of non-human animals and ecosystems to exist and flourish.

While intervention responsibility seems to fulfil the obligations generated by a great number of generators that are very important in our moral landscape, it leaves out some equally important generators of obligation unaddressed. Intervention responsibility is a forward-looking notion of responsibility that aims to mitigate the present harms or prevent future harms, and its scope is quite promising, insofar as the forward-looking obligations are concerned. But I have not yet given any consideration to the agents who were complicit in bringing harm (e.g., high concentration of GHG in the atmosphere) in this model of intervention responsibility. Intervention responsibility as developed so far, notwithstanding its promising scope, does not fulfil obligations generated by some crucial generators which can't be ignored for a comprehensive understanding of responsibility for global warming. In particular, it does not answer the obligation generated by retrospective justice. Retrospective justice focuses on assigning responsibility for past harms and seeking reparations or

compensation for those who have been affected. It involves holding those responsible for past emissions accountable for the harm they have caused and seeking compensation or restitution for communities that have already suffered the impacts of climate change. This could involve legal action, such as lawsuits against fossil fuel companies or governments, or international agreements to provide financial assistance to affected communities. A purely forward-looking notion of responsibility like intervention responsibility focuses on the actions we take in the present and future to address climate change, without taking into account the historical actions and decisions that have led to the current state of the environment. This view of responsibility is insufficient in the context of climate change because it overlooks the significant role that past actions and decisions have played in creating the current situation. Human actions like the combustion of fossil fuels, industrialization, and deforestation have led to the release of greenhouse gases into the atmosphere, causing global warming and ultimately resulting in climate change.

Many of these activities began centuries ago and were carried out by developed countries, companies, or communities that have benefited from them economically. Therefore, one strong criticism of a purely forward-looking notion of responsibility like intervention responsibility is that it fails to acknowledge the historical injustices that have contributed to climate change and unfairly places the burden of mitigating the effects of climate change on agents (individuals or collectives such as companies, countries, etc.) who were not complicit in bringing about the harm. A backward-looking notion of responsibility, on the other hand, recognizes the historical actions and decisions that have contributed to climate change and seeks to hold those responsible for these actions accountable. This approach acknowledges the past and present inequalities in the distribution of responsibility for climate change and

seeks to address them in a more comprehensive way. By looking at the past, we can identify the historical injustices that have contributed to climate change and work towards creating a more just and equitable future.

While it is clear that we need a more comprehensive approach to responsibility ascription in climate change to address various ethical dimensions of the problem, the question that might be raised here is; should we treat backward-looking responsibility and forward-looking responsibility as two mutually exclusive notions of responsibility both of which are necessary for a just distribution of responsibility that make sure that our planet becomes more just and equitable? Or is there a way to integrate these two apparently unrelated notions of responsibility? Or a question that can be specifically targeted at my thesis is; is there a way to incorporate historical responsibility in my notion of intervention responsibility? To the first question, my response is that the distinction between forward-looking and backward-looking notions of responsibility is legitimate and has conceptual as well as normative import. Responsibility ascription is a complex practice and distinguishing forward-looking and backward-looking responsibility allows us to better understand the ethical dimensions of the problem, and to develop a more nuanced understanding of the moral obligations that arise.

That said, one way in which we can potentially integrate historical responsibility into intervention responsibility would be to argue that those who have historically contributed to the climate crises have the historical responsibility to help those, who have relatively contributed less to the crises historically, to develop the intervention strength. This might involve capacity building through various means such as investing in clean energy technologies and supporting adaptation measures etc.

Chapter 6
Identification of Intervention Agents

Intervention responsibility, as formulated in the preceding chapters and understood in the most general sense, is, in principle, applicable to any intervention agent possessing the required intervention strength, regardless of the generators. As explained, intervention agency and intervention strength are graded notions and, thus, differ from agent to agent. Some agents exhibit stronger intervention agency than others, and similarly, some agents show more intervention strength than others. There can be several ways in which different levels of intervention agency and intervention strength can be present in agents, which must be taken into consideration for the ascription of intervention responsibility. The magnitude of intervention responsibility differs among agents having differing levels of intervention agency and intervention strength. Below is a table meant to explain roughly the magnitudes of IR that are ascribable to agents with different levels of intervention agency (IA) and intervention strength (IS). The aggregate measure of both IA and IS will be called intervention capability (IC). Intervention responsibility (IR) increases as IC increases, although other normative factors, such as fairness, cost-benefit ratio, etc. also influence IR. The table below assumes that IC is directly proportional to both IA and IS and that IA and IS are mutually exclusive.

IA/IS	Absent	Partial	Full
Absent	0	0	0
Partial	0	X	Y
Full	0	Z	1

X, Y, and Z denote arbitrary values of intervention capacity, such that X, Y, Z ∈ (0,1)

To illustrate the above table with a practical example, consider a child drowning in the river. Let us start with a single-intervener case where only one person is witnessing the drowning child. An agent bears zero Intervention Responsibility (IR) if it lacks any Intervention Agency (IA) and Intervention Strength (IS). The IR borne by the agent increases along both the axes of IA and IS. If the agent cannot represent the causal model of the situation (IA=0) or does not have any strength (IS=0), the IR borne by that agent is 0. If it can represent the complete model and has the full necessary intervention strength, then it bears an IR of 1. In this case, the agent can make complete sense of what is happening and what can be done to save the child. Therefore, it is morally obligated to intervene by itself. If an agent has an IA of 1 but an IS of 0, what is the responsibility of the agent? Here, the agent is responsible for calling for help and making sure that rescue takes place.

However, in a multi-agent scenario where multiple agents are present, it gets a little trickier and the solutions sometimes sound spooky. Consider that there are multiple agents, each with an IA of 1 but IS that lie between 0 and 1. The agents are fully capable of representing the causal structure of the situation. However, none of them has the strength to intervene individually. Who is responsible to intervene here? In this case, the responsibility of each individual is to join hands and intervene collectively (Hindricks, 2021). The IR in this case is borne by the collective because only the collective exhibits the necessary IS. In the example above, if all the individuals with more or less equal IA and IS, but none with the necessary IS or IA to intervene individually, decide to intervene individually by jumping into the river, it will be a disaster. Of course, the first thing they should do is call for help. But what if there is no possibility of any help coming? Should the individual agents jump into the river and potentially kill themselves in the attempt to rescue the child, or should they just watch from the shore

and let the child die? The best solution for the dilemma is to join hands together and carry out the intervention collectively. Individual intervention is insignificant or futile in this case. However, if multiple agents form a collective, then the probability of obtaining the desired outcome increases.

This example is important for our purpose because it is similar in essence to a broader set of dilemmas generally known as the tragedy of the commons, typical examples of which are deforestation, over-grazing, and global warming. The tragedy of the commons arises in cases where individual contribution is insignificant, but the overall outcome of everyone's contributions is significant. The tragedy of the commons is a term used to describe a situation in which individuals or groups overuse or exploit a shared resource, often resulting in the depletion or destruction of the resource. It is also called the problem of many hands. The problem of many hands is a moral dilemma that arises when a group of people collectively causes some harm, but none of the individuals can be held responsible for it.

While the problem of many hands pertains to a backward-looking dilemma, Hindricks (2021) endorses a different kind of tragedy of the commons called the problem of insignificant hands. It usually arises when an obligation to produce a positive outcome is generated, but individual contributions are insignificant and collective contribution is significant to bring about the outcome. In such cases, it is very difficult to ascribe responsibility to individuals whose potential contribution is futile. For instance, if a person turns off their AC at night, it will not have any significant impact on the overall atmospheric GHG. In such cases, it seems appealing to think that the person has no responsibility to turn off their AC because it will not have any meaningful impact on the outcome that we desire. But it also seems counterintuitive to completely exempt that person from having any responsibility. To solve this dilemma, Walter

Sinnott-Armstrong (Sinnott-Armstrong, 2010) defends the view with a similar example involving joy-guzzlers on a Sunday morning drive. He argues that the individual is not obligated to refrain from going on a drive because their contribution is insignificant and futile. Their responsibility at best would be to hold the government responsible for bringing systematic changes such as banning the use of non-renewable energy in motor vehicles. Shelly Kagan (Kagan 2011) takes the opposite position and argues that if it is, on the whole, more useful to reduce individual emissions of GHG than not reducing them, then individuals must contribute to reducing them.

Hindricks (Hindricks, 2022) defends a moderate solution by proposing an account that he calls the 'prospect account'. On this account, the responsibility to contribute to a positive outcome depends on the feasibility of the contribution. If it is feasible, within certain thresholds, then individuals must contribute to the outcome. In the case of insignificant hands, individuals must come together and contribute collectively so that their collective contribution is feasible to produce the desired outcomes. So, the responsibility falls on the collective in such cases and then trickles down to individuals. The agents here bear responsibility as members of the collective, which has the intervention strength to successfully intervene, but not individually.

The question that can immediately be raised is to ask if a collective can bear any responsibility qua group. Some people might even deny the validity of collectives as legitimate entities possessing intervention agency and intervention strength and argue that only individuals possess intervention agency and intervention strength. The argument that a collective is nothing more than the sum of individuals who constitute it isn't new. It has been one of the two major competing positions about the nature of sociality in the philosophy of social sciences since the 19th century (Smiley, 2022). My response is to take a functionalist stance

and say that groups behave as if they possess the properties often seen as individual agents. I'll further argue that the agency possessed by a collective can be greater than the agencies of individuals who constitute the group.

But before actually taking these solutions to the problem of insignificant hands at face value, a justification of collective agency must be provided. Although I have discussed collective responsibility in the literature review, an account of collective intervention responsibility (CIR), collective intervention strength (CIS), and collective intervention agency (CIA), and how to think of these notions in the interventionist framework, are yet to be provided. In Chapter 4, I argued that agents must be viewed inside a causal model as variables (referred to as agential variables) that have the efficacy to interact with each other and with other non-agential variables inside the causal model. It is easier to imagine individual agents as variables inside the model, but it is not so clear how to view collectives as variables possessing various properties, such as IA and IS, that the individual agential variables inside the causal model possess. In the next section, I'll introduce a rigorous way to find the agents (whether individual or collectives) that have the intervention strength and intervention agency and the causal model that represents these agents and their interactions with other parts of the system.

Coarse-grain modelling

A system, whether natural or social, has a particular causal structure, but it can be described at multiple levels of granularity using different causal models. The most detailed description of the system is called a fine-grained representation, and it describes the system in terms of its micro-states or constituents. In contrast, the description of a higher scale that describes the system in terms of its macro-states or constituents is called the coarse-grained

representation of the system (Hoel, 2017). Coarse-graining refers to the process of simplifying a complex causal system by grouping variables and events into larger, more abstract entities. This allows for a more tractable analysis of the system's behaviour and can help to reveal important patterns and relationships that might not be apparent at a more detailed level. The basic idea behind coarse-graining is to reduce the number of variables and states that need to be considered to describe the system's behaviour. This is done by identifying groups of variables or states that can be treated as a single entity or a "coarse-grained variable." For example, in a physical system, one might group all the particles within a certain distance of each other and treat that group as a single entity with an average position, momentum, and other properties.

In the debate between physicalism and emergentism, coarse-graining serves as a powerful tool that provides empirical verification for causal emergence. Causal emergence is the thesis that says a higher-level description can be sometimes more meaningful and informative than a lower-level description of the system. That is to say, when the macro-states are more efficacious than the micro-states (Hoel et al. 2016). Causal emergence further provides a general explanation as to why different empirical fields of inquiry take different spatiotemporal scales of description for their analyses and even consider those scales to be independent systems (Hoel 2017). For example, biological models ignore nuclear or even atomic interactions for the most part while explaining a biological phenomenon. Erik Hoel (Hoel 2017; Hoel et al. 2016) grounds the phenomenon of causal emergence on another concept from information theory called channel capacity. Taking causal relations as analogous to information channels, he builds a rigorous way to find the causal model and level of granularity that best describes the system. In this formulation, a system has a certain capacity, referred to as the causal capacity of the system, analogous to channel capacity, and the task

is to find the causal model that uses the most of this capacity. This capacity is the ability of the causal structure to convert interventions into unique effects.

Causal emergence using causal models is demonstrated by showing quantitatively that the macro-scale causal models are more informative (and their variables more efficacious) than the underlying micro-models(Hoel et al. 2016). One such measure is the notion of effectiveness, which is a measure of how successful the causal structure is in turning interventions into unique effects. It is further decomposed into two factors: determinism and degeneracy. Effectiveness depends on these factors in the following way: Effectiveness = [determinism] - degeneracy. Determinism is the sufficiency of a given intervention to produce the effects (it also measures how noisy the state-to-state transitions are; the noisier the transitions are, the less deterministic they are). It is a measure of how reliably the interventions produce effects. Degeneracy is the measure of the necessity of intervention; that is, how often transitions result in the same effects. The more degenerate the transition is, the less necessary the interventions are in producing effects. If effectiveness = 1, then the causal relationship between the two states is completely sufficient and necessary (Hoel, 2017).

Formulated in terms of channels, a deterministic causal relation corrects the noise in the channel and converts the inputs into outputs more reliably. Causal emergence happens when the macro-scale causal model exploits the causal capacity of the system more than the micro-scale causal models. The coarse-grained variables are causally more efficacious and their relations are more effective compared to the fine-grained variables.

Coarse-graining of intervention agents

In chapter 4, I introduced the framework of causal modelling and proposed that our system of interest must be extended to include agents as its parts too. Accordingly, the causal model that is supposed to represent it must also represent the agents as variables, occupying particular nodes in the causal network. In this chapter, I began by pointing out the problem of individual responsibility in climate crises and, following others, argued that climate crises are an instance of the problem of many hands, and the most effective way to tackle such crises is to act collectively. Such collective actions require the ascription of collective obligations, which further requires the establishment of collective properties like collective agency and collective strength for their grounding. I concluded the section with a few questions regarding the nature of agency and agents inside the causal models.

In the previous section, I introduced a very powerful technique for finding the entities and their interactions that best explain the emergent phenomena that we see or wish to see. Collective agency is similarly an emergent property that can't be reduced to individual agency, and a collective agent is more than the sum of the individual agents that constitute it (Epstein, 2015). Collective agents must be treated as different entities in themselves. In this section, I'm using the tools from coarse-grained modelling to elucidate how collective agents can be identified within a model. The primary task is to account for why collective agents are more responsible in certain cases, like in the problem of many hands, by using the interventionist framework to provide some descriptive grounds, as I did in the previous chapter for individual responsibility.

In our extended system, we have two kinds of variables: the agential variables and the non-agential variables. Both of these kinds of variables are causally efficacious and interact with each other. The non-agential variable may be the emission of greenhouse gases into the air. But the emissions that contribute majorly to the net greenhouse gas concentration are not the minute emissions caused by individuals in their daily lives, but rather the huge amount of emissions (collective emissions) caused by various collective entities, such as corporations, populations, etc. These larger quantities of greenhouse gases are basically what explain the increment in the net greenhouse gas and consequently the warming of the mean surface temperature of the earth, known as global warming. Similarly, as discussed at the beginning of this chapter, the entities that have the best chance of tackling crises like climate change aren't the individual agents but the collective agents. So, the extended system does not just contain a large number of individual greenhouse gas molecules (or even contributions from individual agents in the form of smaller traces) and individual agents as its variables. But collective agents and collective emissions are parts of the system too. And it is these parts that are causally more efficacious and account for the warming of the earth as well as potential action against the crises.

Coarse graining is the way to reach these bigger entities and their interactions with the basic constituents (individual agents and individual emissions) and their interactions. Here, I'm particularly concerned with the coarse-graining of the agential variables. The task is to identify the causal model and granularity of description at which the agential variables are most effective in influencing the coarse-grained non-agential variables to produce the desired manipulation. So, the causal model that we need to construct is not descriptive, in the sense that it will just represent entities (and their interactions) that already exist, but it will represent entities that can be constructed and which can potentially alter the system as a whole.

The first step in the process of coarse-graining intervention agents is to identify the agents in the system that you want to coarse-grain. This could involve grouping individuals into smaller or larger groups. The choice of agents will depend on the skills, resources, and knowledge possessed by the agents. For instance, to have a strong collective agent that can tackle the climate crisis in a particular region, we need people with knowledge of climate science, technology, economics, sociology, and ethics, as well as non-academic persons like mechanics, technicians, etc. A collective having agents with these complementary skills will have a better chance of taking concrete action than, say, a collection of climate scientists. The choice of selecting particular agents for coarse-graining is based on the condition that the coarse-grained agent must have the necessary strength for intervention.

Once the intervention agents are identified, the next step is to define the level of abstraction at which you want to coarse-grain them. In this step, the task is to find the size of the coarse-grained agent. After defining the level of abstraction, the next step is to determine the causal relationships between the coarse-grained intervention agents and other components of the model. This could involve identifying direct causal connections between the coarse-grained agents, as well as indirect causal effects through other agents. This step is critical for ensuring that the coarse-grained model accurately captures the dynamics of the system and that potential alterations can be brought.

I have so far argued that coarse-graining is a way to find coarse-grained agents in a system, and the coarse-grained agents, which are collective agents, are more effective at tackling the issue of GHG emissions. But I have not yet provided a descriptive explanation of the effectiveness of these coarse-grained agents. For this purpose, we need to construct a causal

model that represents these coarse-grained agents and their interactions with the non-agential variables (as well as with each other).

As discussed in the previous section, certain emergent phenomena can be better explained through causal emergence rather than by taking a physicalist approach. That is to say, certain emergent phenomena can be better explained in terms of the interactions of some coarse-grained variables (or macro-states) rather than in terms of fine-grained variables (or micro-states). This happens when the interactions between coarse-grained variables are stronger, and the causal relationships more robust. The interventions carried out on those coarse-grained variables produce the relevant effects more reliably. Borrowing from this, we can say that coarse-grained agents are more effective in carrying out interventions. This effectiveness is the measure of the strength of intervention agents.

The effectiveness of the intervention agent in carrying out the intervention depends on two factors: determinism and degeneracy. Different causal models representing a particular system can have causal relations with differing levels of determinism and degeneracy. In the case of climate crises, formulated as a problem of insignificant hands, the causal relations between individual hands and the desired outcome are not deterministic and hence insufficient. Individual contributions are utterly insignificant in producing the outcomes. However, collective contributions can be significant, and thus, the causal relationships between coarse-grained agents and the variable of interest are more deterministic. They produce the outcome (reduction in GHG emission) more reliably.

Within the macro-causal model, coarse-grained agents are causally more efficacious. Similarly, the causal relationships between the individual agents and the variable of interest are more degenerate because, oftentimes, many individuals contribute to just one of the many parts of

the problem, leaving other parts of the problem unaddressed or unsolved. When tackling the whole problem, it requires distributing labours to every part of the problem. The hands should contribute to specific parts of the problem to produce an outcome. But what happens in the case of insignificant hands is that multiple hands dig the same spot of the problem to produce an insignificant outcome.

Therefore, a macro-scale description of the system describes the system more accurately, and it has great utility for the ascription of intervention responsibility too. The coarse-grained intervention agents are causally more efficacious; the relationships between the coarse-grained agents and other variables are more robust and have more effectiveness compared to the relationships between fine-grained intervention agents (individual agents). If we assume causal relationships in the model to be analogous to information channels, then the coarse-grained agents use more of the channel capacity. They are more likely to correct the errors in the channel that otherwise obtain in the channels between the fine-grained variables.

Coarse-grained agents have more intervention strength because their effectiveness in carrying out an intervention is greater than the individual agents, and the values assumed by them are more sensitive to the causal structure than the values assumed by the individuals. To put it intuitively, collective agents are more able to channelize their contributions than individual agents to produce the desired outcome. In conclusion, I argue that the problem of insignificant hands is actually a problem of causal model choice, and it arises because a wrong causal model is assumed when looking at the ways to change the system. In particular, it arises because we choose a causal model at the micro-scale of the description that consists of fine-grained variables, instead of a macro-scale consisting of coarse-grained variables.

Collective intervention agency of coarse-grained agents

I have so far given an account of the intervention strength of the coarse-grained agents within a system represented by a causal model. The intervention strength of the collective agents, like that of individual agents, is possessed by the collective but conditioned by the causal structure of the system. But one crucial aspect of the coarse-grained agents is not yet discussed. For the most part in this chapter, I have referred to the agents in the system as variables, entities, and states. But we know that agents are not completely similar to other variables like GHG emissions, etc. Agential variables have a different ontological status than non-agential variables. They are different by the virtue of having intervention agencies. Collective agents, or coarse-grained agents, have the tendency to represent causal structures through causal models too. Additionally, the intervention agency possessed by coarse-grained agents may be greater than the intervention agency of fine-grained agents. Collective agents may have a greater tendency to represent a causal structure than individuals for several reasons, such as:

Division of labour: Collective agents are composed of multiple individuals who have specific roles and responsibilities. This division of labour allows for greater specialization and expertise in specific areas. As a result, collective agents are better equipped to understand and represent complex causal structures that may involve multiple factors and variables. For example, a research institute that studies climate change may have scientists, economists, and policymakers each of whom brings a unique perspective and expertise to the problem. Together, they can develop a more comprehensive understanding of the causal factors that contribute to climate change and propose effective solutions. With a comprehensive model, effective interventions can be represented by the group.

Group decision-making: Collective agents often make decisions through a process of group deliberation and consensus-building. This process allows for the consideration of multiple viewpoints and the synthesis of diverse perspectives. As a result, collective agents are more likely to represent a causal structure more comprehensively and reliably.

As with individual intervention responsibility, collective intervention responsibility is ascribed to a coarse-grained entity once the collective intervention agency and the collective intervention strength are established.

Chapter 7
The Social Ontology of the Coarse-Grained Agent

In the previous chapter, I argued that the type of agent with the best chance of making an effective and significant contribution to mitigating the climate crisis is the coarse-grained or collective agent. I then provided a technique to find coarse-grained agents in a system through a causal model and gave a descriptive justification as to why those coarse-grained agents are more responsible than the fine-grained agents in the system. However, I left several crucial issues regarding collective agents and their properties unaddressed. Three main problems are worth discussing, which I will attempt to address in this section.

Firstly, coarse-graining was proposed as a way of finding coarse-grained or collective agents, but it was not clear what criteria make a collection of individuals a collective agent. Grouping some random people in an arbitrary way will not necessarily result in a collective agent. It is only certain collections, known as collective agents, that have the collective agency and intervention strength necessary to intervene. Once we have a general framework to characterize various group agents, it will be easier to identify and classify various groups according to their strengths. With this knowledge, we can also potentially create group agents that possess a certain degree of intervention strength.

Secondly, the relationship between fine-grained agents and coarse-grained agents was not addressed, particularly the metaphysical dependence of collective agency on individual agency. This issue is crucial because it can illuminate different ways to construct group agency and different ways in which collective responsibility can be distributed among group members. The field of inquiry that deals with the nature of social entities such as group agents

is called social ontology, and more broadly social metaphysics. It is concerned with various entities in the world that emerge from social interaction (Epstein, 2021). Although there are disagreements about what exactly a social group is--whether it is a class, fusion, set, structure, or some other kind of entity-I take the view that social groups are distinct from their members and they not only supervene on its members but also on other features and factors (Epstein, 2015).

The question then arises: what makes social groups more than the collection of people that constitute them? As the name suggests, there is a social part to the group that a random collection of people presumably does not possess. The issue here reduces to the nature of sociality. Some people have taken this approach by arguing that the issue of group agency is basically a question about sociality. It has generally been argued that in social groups, members have collective attitudes, shared intentions, and joint commitments. These features are what create sociality in a group of individuals. Different authors have stressed features like joint commitment, collective attitudes, and shared intention to different degrees as legitimate generators of sociality. Notable contributors to this tradition are Margaret Gilbert (Gilbert, 1989), who endorses the joint commitment approach; Michael Bratman (Bratman, 2014), who endorses shared intention; and Raimo Toumela (Tuomela, 2013), who proposes the collective attitude (mode-account) model of group agency. They all argue that those groups that have one of these internal structures that integrate them essentially have group agency.

Another influential theory of social groups is due to Katherine Ritchie, who classifies social groups into two categories: organized groups and feature groups (Ritchie, 2015). Organized groups have certain structures consisting of nodes and hierarchies, and these roles are taken

up by replaceable people. These groups can have collective intentionality. The second category is ‘feature groups’, which are unstructured and incapable of forming collective intentions.

In contrast to other approaches, Brian Epstein (Epstein, 2015) believes that group agency exhibits heterogeneity in grounding conditions, and different groups manifest group agency in diverse ways on different grounds. Epstein argues that group agency is a functional property and is multiply realisable; it is realised differently in different groups. He refutes the notion that integration of members (attitudes, intentions, commitments, etc.) is a necessary or sufficient condition for a collection to be a group agent (Epstein, 2015). Epstein (Epstein, 2017) rejects the approach that reduces agency to sociality, stating that not every social group is a group agent and that group agents are only a subset of the broader category of social groups.

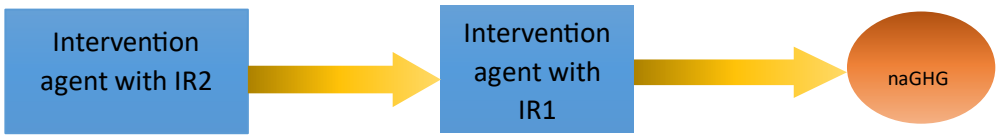
Even if we accept that there is no singular criterion that grants the emergence of group agency from a collection of individuals, we can still appreciate the importance of various criteria, such as the presence of rigid structures, joint commitments among members, collective attitudes, collective representations, reasoning, etc. With these criteria in place, we have a way to identify and create group agents that can represent certain properties and play the functional role of agency.

Regarding the ontological relationship between individual agency and group agency, some researchers consider it to be supervenience. It is apparent that group agents comprise mainly of people, but group agency is an emergent phenomenon that cannot be reduced to individual agency. The emergence of group agency results from the integration of various individuals' mental states and the structure of which the individuals are a part. The collective

intervention responsibility assigned to a group agent is distributed among individuals of that group according to their roles. If the group lacks a strong structure, the responsibility that the group can be assigned also decreases. In ascribing intervention responsibility, a critical task is to identify those groups that have strong organisational structures and the ability to deliberate and reason collectively by representing causal models and interventions.

Along with identification, the creation of group agents is also crucial. Individuals in the system who lack the intervention strength or intervention agency required to make a significant contribution to the mitigation of climate crises have a responsibility to create group agents. The process of collectivisation of agency gives rise to group agency. Once a group is created and starts exhibiting an emergent group agency, it can then potentially possess the collective intervention strength and agency necessary to make a significant contribution. Individuals who lack the required intervention strength and agency have the responsibility to hold other capable agents responsible. However, this may not always be feasible. Ordinary individuals are generally unable to force influential agents (whether collective or individuals) to take morally right actions. But through the collectivisation of their individual agencies by forming organised bodies, they can gain power to force those influential agents to take morally desirable actions, such as climate actions (Hindriks, 2019)(Collins 2013). The collectivisation of individual agents into a collective agent, thereby forging the individual agencies into a group agency, may involve several steps. First, a certain number of people must be selected based on their individual strengths, expertise, and knowledge. Once selected, there must be a joint commitment among the members to achieve the common goal. The selected individuals must also have a collective attitude towards the task by thinking about the problem as a collective problem and solving it as a collective burden. The next step may involve setting up a formal organisation with a rigid structure that assigns roles to different

people. Finally, individuals with different skills, knowledge, and expertise can be assigned roles within that organisation. Once such a group agent is constructed, it gains strength and collective agency and becomes a proper site of intervention responsibility. However, their intervention responsibility primarily amounts to holding the primary bearers of intervention responsibility, such as the United Nations or big oil mining companies, responsible for intervening. Understood within the causal modelling framework, their responsibility is to intervene with other agents and induce them to take actions that they are obligated to. It has also been suggested that the ascription of responsibility itself must be thought of as an intervention (Hitchcock & Knobe, 2009). In this sense, both the assigner and the assignee of intervention responsibility are agential variables inside the system. The assignee usually bears the primary intervention responsibility of reducing the emission of net atmospheric GHG by intervening directly on the emission variable. Let's call this responsibility IR1. The assigner of IR1 bears the intervention responsibility of forcing other, more capable agents to take necessary actions by intervening on those agential variables. Let's call this responsibility IR2. The degree of IR trickles down from highly organized and capable group agents to less organized group agents to individuals.



Conclusion

In conclusion, this thesis has provided a rigorous and coherent framework for understanding moral responsibility in the context of climate change. After reviewing the existing literature on responsibility and its different variants, I argued that a forward-looking approach to moral responsibility, which takes collectives as primary bearers, is the most appropriate for addressing the challenges posed by climate change. Moreover, I have argued that we need to develop a more sophisticated model of forward-looking responsibility compared to the existing ones, to adequately account for the complexity and systematicity of the problem.

To ground this approach, I drew on the interventionist framework of causation that allows us to understand how our actions can intervene in the complex causal relationships underlying climate change. By using this framework, we can better understand our individual and collective responsibilities to mitigate the effects of climate change. Therefore, I have provided a novel and robust grounding to a more sophisticated model of forward-looking responsibility known as intervention responsibility. In particular, I have proposed a new kind of agency known as an intervention agency that is necessary for the ascription of intervention responsibility. I have given an account of intervention strength using the resources from the interventionist theory of causation as well as causal emergence. In addition, I argued that coarse-grained modelling can provide a useful tool for understanding the ways in which collective agents have more capacity to intervene in these causal relationships than individual agents. By assessing the intervention capacity of different types of agents in a system using causal models, we can fairly allocate responsibility to various actors in a collective action problem of this scale and complexity. Furthermore, I have argued that different kinds of justice

work as generators of responsibility rather than as grounds, and it is equally important to reveal the generators as much as it is to provide grounds.

Moving forward, there is much work to be done in addressing the challenges posed by climate change. While my thesis provides a useful framework for understanding moral responsibility, a great deal of work can be done to take this framework further. Firstly, the model can be made more rigorous, potentially by formalising notions like intervention agency and strength further and finding a way to quantitatively evaluate intervention responsibility. Secondly, the normative aspect of both intervention responsibility as a moral status and the practice of the ascription of responsibility needs to be further explored. This thesis has primarily been concerned with providing a fairly objective framework and ground for intervention responsibility, which is needed for any notion of responsibility. Furthermore, I have not even touched on the application of intervention responsibility. The question of which specific entities, such as nations, corporations, or organisations, are actually intervention responsible to intervene in climate change is a purely applied one, and I have not delved into it. There is literature on the responsibility of corporations, nations, etc. in general, as well as their historical responsibility for emissions. One can draw from these areas to assess the intervention responsibility of such entities.

Ultimately, it is my hope that this thesis will contribute to the growing yet still nascent field of climate ethics. By recognizing our individual and collective responsibilities to mitigate its effects, we can work towards a more sustainable and just future for all.

References

Bratman, Michael E. 2014. *Shared Agency: A Planning Theory of Acting Together*. OUP USA.

Chalmers, David. n.d. "What Is Conceptual Engineering and What Should It Be?" *Inquiry: An Interdisciplinary Journal of Philosophy*. doi: 10.1080/0020174x.2020.1817141.

Collins, Stephanie. 2013. "Collectives' Duties and Collectivization Duties." *Australasian Journal of Philosophy* 91(2):231–48.

Copp, David. 2010. "Corrective Justice as a Duty of the Political Community: David Lyons on the Moral Legacy of Slavery and Jim Crow." *Boston University Law Review. Boston University. School of Law* 90.

Epstein, Brian. 2015. *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. Oxford, New York: Oxford University Press.

Epstein, Brian. 2019. "What Are Social Groups? Their Metaphysics and How to Classify Them." *Synthese* 196(12):489–493.

Epstein, Brian. 2021. "Social Ontology." in *The Stanford Encyclopedia of Philosophy*.

Fabre, Cécile. 2012. "Humanitarian Intervention." P. 0 in *Cosmopolitan War*, edited by C. Fabre. Oxford University Press.

Fabre, Cécile. 2014. *Cosmopolitan War*. Oxford, New York: Oxford University Press.

French, Peter A. 1984. *Collective and Corporate Responsibility*. New York Chichester, West Sussex: Columbia University Press.

Gardiner, Stephen M. 2011a. *A Perfect Moral Storm: The Ethical Tragedy of Climate Change*. Oxford University Press, USA.

Gilbert, Daniel. 2006. "If Only Gay Sex Caused Global Warming." *Los Angeles Times*. Retrieved April 1, 2023 (https://www.latimes.com/archives/la-xpm-2006-jul-02-op-gilbert2-story.html).

Gilbert, Margaret. 1989. *On Social Facts*. Routledge.

Grinfeld, Guy, David Lagnado, Tobias Gerstenberg, James F. Woodward, and Marius Usher. 2020. "Causal Responsibility and Robust Causation." *Frontiers in Psychology* 11:1069. doi: 10.3389/fpsyg.2020.01069.

Halpern, Joseph Y., and Christopher Hitchcock. 2015. "Graded Causation and Defaults." *British Journal for the Philosophy of Science* 66(2):413–57.

Hehir, Aidan. 2012. *The Responsibility to Protect: Rhetoric, Reality and the Future of Humanitarian Intervention*. Bloomsbury Publishing.

Hindriks, Frank. 2021. “The Problem of Insignificant Hands.” *Philosophical Studies* 179(3):829–54.

Hitchcock, Christopher, and Joshua Knobe. 2009. “Cause and Norm.” *Journal of Philosophy* 106(11):587–612. doi: 10.5840/jphil20091061128.

Hjorthen, Fredrik D., and Göran Duus-Otterström. 2016. “Humanitarian Intervention and Historical Responsibility.” *Journal of Global Ethics* 12(2):187–203. doi: 10.1080/17449626.2016.1194308.

Hoel, Erik P. 2017. “When the Map Is Better than the Territory.” *Entropy* 19(5):188. doi: 10.3390/e19050188.

Hoel, Erik P., Larissa Albantakis, William Marshall, and Giulio Tononi. 2016. “Can the Macro Beat the Micro? Integrated Information across Spatiotemporal Scales.” *Neuroscience of Consciousness* 2016(1):niw012. doi: 10.1093/nc/niw012.

Icard, Thomas F., Jonathan F. Kominsky, and Joshua Knobe. 2017a. “Normality and Actual Causal Strength.” *Cognition* 161:80–93. doi: 10.1016/j.cognition.2017.01.010.

Icard, Thomas F., Jonathan F. Kominsky, and Joshua Knobe. 2017b. “Normality and Actual Causal Strength.” *Cognition* 161(C):80–93.

Isaacs, Tracy. 2014. “Collective Responsibility and Collective Obligation.” *Midwest Studies in Philosophy* 38(1):40–57.

Jacob, Pierre. 2003. “Intentionality.” in *Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta.

Jamieson, Dale. 2007. “When Utilitarians Should Be Virtue Theorists.” *Utilitas* 19(2):160.

Jamieson, Dale. 2014. *Reason in a Dark Time: Why the Struggle Against Climate Change Failed -- and What It Means for Our Future*. Oxford University Press.

Jamieson, Dale. 2015. “Responsibility and Climate Change.” *Global Justice: Theory Practice Rhetoric* 8(2).

Jones, Matthew W., Glen P. Peters, Thomas Gasser, Robbie M. Andrew, Clemens Schwingshackl, Johannes Gütschow, Richard A. Houghton, Pierre Friedlingstein, Julia Pongratz, and Corinne Le Quéré. 2023. “National Contributions to Climate Change Due to Historical Emissions of Carbon Dioxide, Methane, and Nitrous Oxide since 1850.” *Scientific Data* 10(1):155. doi: 10.1038/s41597-023-02041-1.

Kagan, Shelly. 2011. “Do I Make a Difference?” *Philosophy and Public Affairs* 39(2):105–41.

Lagnado, David A., Tobias Gerstenberg, and Ro’i Zultan. 2013. “Causal Responsibility and Counterfactuals.” *Cognitive Science* 37(6):1036–73. doi: 10.1111/cogs.12054.

Lilli, Eugenio. 2019. “Responsibility to Protect (R2P).” Pp. 1–5 in *The Palgrave Encyclopedia of Global Security Studies*, edited by S. Romaniuk, M. Thapa, and P. Marton. Cham: Springer International Publishing.

Luck, Edward C. 2011. “The Responsibility to Protect: The First Decade.” *Global Responsibility to Protect* 3(4):387–99. doi: 10.1163/187598411X603025.

Lyons, David. 2003. “Corrective Justice, Equal Opportunity, and the Legacy of Slavery and Jim Crow.”

Menzies, Peter, and Huw Price. 1993. “Causation as a Secondary Quality.” *British Journal for the Philosophy of Science* 44(2):187–203.

Mill, John Stuart. 2020. “A Few Words on Non-Intervention.” *Sotsiologicheskoe Obozrenie / Russian Sociological Review*.

Nussbaum, Martha C. 2006. *Frontiers of Justice: Disability, Nationality, Species Membership*. Harvard University Press.

Pattison, James. 2013. “Is There a Duty to Intervene? Intervention and the Responsibility to Protect.” *Philosophy Compass* 8(6):570–79.

Pearl, Judea. 2000. *Causality*. Cambridge University Press.

Pogge, Thomas W. 1992. “An Institutional Approach to Humanitarian Intervention.” *Public Affairs Quarterly* 6(1):89–103.

Rawls, John. 1993. “The Law of Peoples.” *Critical Inquiry* 20(1):36–68. doi: 10.1086/448700.

Ritchie, Katherine. 2015. “The Metaphysics of Social Groups.” *Philosophy Compass* 10(5):310–21.

Roth, Abraham Sesshu. 2011. “Shared Agency.” in *Stanford Encyclopedia of Philosophy*, edited by E. Zalta.

Samland, Jana, and Michael R. Waldmann. 2016. “How Prescriptive Norms Influence Causal Inferences.” *Cognition* 156(C):164–76.

Schlosser, Markus. 2015. “Agency.” in *Stanford Encyclopedia of Philosophy*.

Sinnott-Armstrong, Walter. 2010. “It’s Not My Fault: Global Warming and Individual Moral Obligations.” P. 0 in *Climate Ethics: Essential Readings*, edited by S. M. Gardiner, S. Caney, D. Jamieson, and H. Shue. Oxford University Press.

Sinnott-Armstrong, Walter. 2019. “Consequentialism.” *Stanford Encyclopedia of Philosophy*.

Sloman, Steven. 2005. *Causal Models: How People Think about the World and Its Alternatives*. Oxford University Press.

Smiley, Marion. 2008. “Collective Responsibility.” in *Stanford Encyclopedia of Philosophy*.

Spirtes, Peter, Clark Glymour, Scheines N, and Richard. 2000. *Causation, Prediction, and Search*. Mit Press: Cambridge.

Tuomela, Raimo. 2013b. *Social Ontology: Collective Intentionality and Group Agents*. Oxford, New York: Oxford University Press.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.