

# Collective Agents as Moral Actors

Säde Hormio

Practical Philosophy, University of Helsinki

Forthcoming in Hormio and Wringer (eds.) *Collective Responsibility: Perspectives on Political Philosophy from Social Ontology*. Studies in the Philosophy of Sociality series.

## Abstract

How should we make sense of praise and blame and other such reactions towards collective agents like governments, universities, or corporations? Collective agents can be appropriate targets for our moral feelings and judgements because they can maintain and express moral positions of their own. Moral agency requires being capable of recognising moral considerations and reasons. It also necessitates the ability to react reflexively to moral matters, i.e. to take into account new moral concerns when they arise. While members of a collective agent are capable of this, the collective frames the thinking of the individual moral agents within it and affects their options in myriad ways. The moral positions thus formed and expressed belong to the collective. Crucially, unlike marginal moral agents, collective agents as moral actors can be held fully responsible for their acts and omissions.

**Keywords:** collective responsibility, collective agency, moral agency, moral sentimentalism, blame.

## 1. Introduction

A stone bench on a university campus bears the description: “In gratitude and loyalty to our alma mater”. In her retirement speech, a conductor remarks about the love she feels for the orchestra she has led for many years. Angry employees gather in the company car park to protest the way their employer is treating them. A prime minister expresses public regret over historical injustices towards the nation’s indigenous community and asks for their forgiveness.

How should we make sense of such reactions towards collectives? Perhaps with the orchestra, we can think of the conductor expressing her aggregate love for all the individual musicians that make up the orchestra. In a similar fashion, maybe the disgruntled employees’ demands are directed at the owner of the company, or the board of trustees. The reactions would then be directed towards individuals as role-occupiers. However, we are missing something crucial if we try to reduce these reactions to an aggregate of sentiments towards individuals. The conductor

expresses her love for the orchestra itself, for how the different musicians come together to form a musical whole. The employees might feel let down by their employer in a way that goes beyond what individual role-occupiers have done or omitted to do. Individualistic explanations are not even available for the other two examples. The class that donated the bench to the university graduated decades earlier and no longer has any personal relations with the role-occupiers at the university (here both the object and the subject are collective).<sup>1</sup> The prime minister is not apologising for harms inflicted by people alive today, but in the name of the nation for injustices committed by previous generations. In all these cases, the reactions are directed at the whole, not towards the parts (although some parts might be held up to special scrutiny).

The blame the employees express is meant to hold the *collective agent* responsible for the way it treats them. Collective agents are organised, structured collectives with a shared goal or purpose, an ethos, and often hierarchical roles. Institutions and corporations are paradigm examples of integrated collective agents. I will argue that collective agents can be appropriate targets for our reactive attitudes – such as blame, praise and gratitude – because they can maintain and express moral positions of their own. Reactive attitudes are part of what forms our relationships: what we demand and expect of other actors in the moral sphere regarding their intentions and attitudes towards us.<sup>2</sup> Moral positions are views about the value of other agents or entities, expressed through words or actions, or both. I will label agents that can maintain and express moral positions of their own as *moral actors*. Collective agents as moral actors are holistic entities towards whom reactive attitudes are sometimes rightly directed.<sup>3</sup>

With the notion of collective responsibility, reactive attitudes are directed at the group level. As a group-based construct, the source of moral responsibility is in the group actions and omissions. Those who criticise collective responsibility as a moral construct acknowledge the convenience of the concept, but express doubts about how the absence of outright collective mental lives impacts it (Smiley 2022). It is perhaps easy to think that maybe we should not expect much from collective agents morally. Some might suggest that we should conceptualise collective agents as

---

<sup>1</sup> It is the class of 1897 who wanted to express their gratitude and loyalty to the university that they formerly attended, although individuals had to be active in the alumni association to make the donation. Maybe the sentiment expressed belonged only to a section of the students from that graduating class, and perhaps we could conceptualise the feelings of gratitude toward a particular time-slice of the institution, rather than the institution as a whole.

<sup>2</sup> P. F. Strawson (1962) famously argued that the human reactions we naturally have when confronted by the good or ill will of others (as displayed in their actions and attitudes) are manifestations of morality.

<sup>3</sup> The default assumption is that to be a legitimate target for reactive attitudes, one needs to be a moral agent. One upshot of my argument is that this can be called into question, although my solution is still anchored in the moral agency of the members of the collective.

something akin to marginal moral agents, for example based on the way they are legally constructed. Legal theorist Joel Bakan (2004) has famously claimed that a modern corporation is a pathological institution. According to him, the way corporations are created by law allows them to focus single-mindedly on profit creation, and this makes them function akin to a psychopathic personality. This might mean that we should demand less of them morally than of other agents, at least until we fix the legal arrangements that underpin modern corporations. After all, moral agency comes in degrees, where marginal moral agents are eligible for some responsibility responses, but not all (Shoemaker 2015). Marginal moral agents can include, for example, people with intellectual disabilities, dementia, pathological conditions (e.g. psychopaths), or those with temporary mental disorders (like severe depression). Even though such agents do not lack moral standing, the fact that we do not expect as much from them as from full moral agents means that we do not evaluate them in the same way (Erskine 2003). Perhaps collective agents are like this too.

I want to resist any such conclusion.<sup>4</sup> Rather than put collective agents in the category of marginal moral agents, or label them as somehow deficient moral agents (Arpaly 2015), I believe that we can hold them fully morally responsible. In other words, while marginal moral agents cannot be held fully responsible, collective agents as moral actors can be.<sup>5</sup> The concept of moral actors can help to explain why our reactions towards collective agents can make sense, without the need to widen the moral agency conditions established within literature on individual moral responsibility.<sup>6</sup> Moral agency requires, among other things, being capable of recognising moral considerations and reasons, as well as the ability to react reflexively to moral matters, by which I mean being able to take into account new moral concerns when they arise. I will call the ability to grasp novel moral information and apply moral reasons to new situations *moral reflexivity*. While members of a collective agent are capable of this, the collective frames the thinking of the individual moral agents within it and affects their options in myriad ways. The moral positions thus formed and expressed belong to the collective. Collective agents as moral actors act with

---

<sup>4</sup> I agree that the set-up and legal requirements of corporations need rethinking to better encourage and recognise their nature as moral actors, and not to give corporations disincentives to act irresponsibly. That is a matter for another paper, however. Here I will concentrate only on how we can and should demand that collective agents take moral concerns seriously and how we can hold them fully morally responsible, as they act with moral valence.

<sup>5</sup> This does not mean that they necessarily are, it just means that they do not somehow automatically disqualify from full responsibility just by virtue of being collective agents.

<sup>6</sup> It might well be that many collectivists agree with what I am arguing, but it can be confusing to widen the category of moral agents to involve collective moral agency, as such a move needs to come with its own set of disclaimers about how it fits with what we are used to viewing as moral agents. Note that this is not to say that the collectivist position needs to be wrong, rather my argument is aimed at those on the fence about the concept of collective responsibility or who want to hold onto individualistic moral agency.

moral valence: they accrue moral responsibility through the acts and omissions that follow from their decisions and values.

I believe that it is helpful to have the separate category of moral actors, as it brings the debates from individual and collective moral debates closer together. It can also help to give shape to some widely shared common intuitions about collective agents acting within the moral sphere. My argument also highlights the importance of individual decisions taken within roles and the collective context. This allows us to appreciate both how individuals can affect the collective and how the collective affects its members. Moral agency requires moral reflexivity, which is connected to moral understanding and feelings through the ability to grasp novel moral information and applying moral reasons to new situations. In the usage that I propose, moral actor is a wider category than moral agent. It includes all agents that can maintain and express their own moral positions, regardless of whether they are moral agents or not. Along with collective agents, artificial intelligence might also come under the moral actor category as it becomes more sophisticated.<sup>7</sup>

The argument proceeds as follows. I begin by setting out some preliminaries about collective agents in section two, before offering my account of how we should understand collective agents as moral actors in section three. As moral actors they are able to maintain and express collective moral positions, which reflect how the collective reacts to something as a collective. In section four, I explore a functionalist account of collective moral agency and its limits.

## **2. Collective agency – preliminary stipulations**

I will next make some preliminary stipulations about the notion of collective agency as background for the arguments to come, to make the chapter accessible also to those without prior knowledge of the literature. Instead of dwelling on the complex topic of agency, I will follow several writers who have commented on it to try to adopt a relatively uncontroversial view of collective agency. A fictional corporation, Atlantis, is introduced to tease out the distinctions introduced. Atlantis is a large online retailer that pays low wages to its warehouse workers while failing to provide safe working conditions. Many people at the warehouses have

---

<sup>7</sup> Carissa Véliz (2021) describes current AI, algorithms and robots as moral zombies: although they can have moral impact on the world, they lack sentience. Stephanie Collins (2022) argues that a present-day robot cannot grasp or feel reactive emotions and lacks awareness of having done something wrong.

had to quit after just a few years due to debilitating back conditions or injuries due to inadequate health and safety measures. Atlantis complies with laws, but spends millions supporting politicians who favour lax regulation, and lobbies against initiatives that would broaden the scope of its responsibilities towards its workers in the eyes of the law, including increasing the minimum wage.

To have agency, an entity must be able to act in the world by choosing between actions and possible outcomes. Mechanisms lack agency: agency cannot be predetermined or random, but rather is an activity that originates within the agent (Riskin 2016, p. 3). A collective should have a structure appropriate for generating intentional actions to qualify as an agent (Isaacs 2014, p. 42). Atlantis is paying its warehouse workers low wages not due to some oversight, but because it has decided to do so: it is acting intentionally in relation to its wage policy and lobbying activities. Highly structured organisations like Atlantis are the clearest examples of collective agents. Their decision-making procedures allow them to act in a coherent manner and to make long-term plans and commitments, be consistent or inconsistent, enter into agreements, keep or break their promises, change their course of action, and so on. They also have a set of core members, such as a board of trustees, that can be identified as representing the collective. Atlantis is a unified collective agent that involves legal properties and material possessions, constitutive features (such as group goals, beliefs, and norms), and individual members (Tuomela and Mäkelä 2016).

While individuals constitute collective agents to an important degree (Collins 2022), collective agents are never just aggregates of individual members. Highly organised large collectives are the easiest to differentiate from any set of individual members (Isaacs 2014, p. 24). Atlantis is not reducible to its CEO, its board of trustees, its workers, or its shareholders. These people all act within the confines of their roles, although they naturally have much more leeway than more peripheral members. Still, what Atlantis decides is not the sum of their individual plans and decisions. The decisions of collective agents can and do differ from the aggregate opinions or choices of individual members (List and Pettit 2011). This kind of collective agency is rational agency and often also legal agency. However, I will not discuss legal aspects of collective responsibility, as a course of action can be legal, yet considered unethical, like using sweatshop labour in countries with lax regulation, or employing novel data-gathering technology in an irresponsible manner. Blame directed at collective agents in such situations not only pre-dates legislation, but it can also act as the pressure that spurs new legislation. Atlantis values customer satisfaction and profit over the wellbeing of its workers. When the employees in the car park

protest about how their employer is treating them, they are targeting their anger at the unjust moral position that Atlantis is expressing, even when it is acting within the law. They are also calling the attention of outsiders to the same by publicly blaming the corporation.<sup>8</sup> The workers in the warehouses are challenging the implicit claim embedded in the actions of Atlantis that it is acceptable to treat them this way, that they are not worth more respect and care (on such a reading on blame, see Hieronymi 2001; Smith 2013; Srinivasan 2018). The collective agent is the legitimate target of the blame and the responsibility claims that follow from the moral position it has taken as a moral actor.<sup>9</sup>

Corporate intentions and actions cannot be reduced to those of individual members partly because corporate actions require an appropriate institutional context and constitutive rules like an ethos – the characteristic spirit and culture of the collective as manifested in its goals and beliefs. Think of a bakery, Chez Louis, that has decided to bring out a new product line. When its managers discuss the pros and cons of widening the product range, the point is not if they as individuals enjoy biscuits. Rather, they look at the matter with their bakery-membership hats on: what are their competitors doing, what the market looks like, how have their customers reacted to sample items, what makes sense in light of the brand, and so on. These debates give rise to a group belief about what is the best course of action for Chez Louis from the point of view of its market position, traditions, ways of working, etc. In doing so, members must consider material things like the resources available, as well as the company's norms around introducing changes. The individual deliberations within a role are always framed by the group, within the parameters of the collective ethos and under its influence. Collective ethos thus frames the individual deliberations. It covers the collective agent's main goals and commitments, i.e. the questions and practical matters vital for it (*realm of concern*) and the answers to these it has accepted as its view (*intentional horizon*) (Tuomela 2007, p. 15; Laitinen 2014). While the ethos determines the collective identity, it is not set in stone. In the case of Chez Louis, the realm of concern continues to be to offer the best possible bakery products at an affordable price-point, while still sourcing ingredients responsibly, but the intentional horizon has been expanded to include making biscuits. Such changes in the ethos are in no way rare or exceptional, as collective agents regularly review and adjust their goals and commitments.

---

<sup>8</sup> The response from the workers is resentment, based on the relation they have to the collective agent. Agents not directly affected by its actions (e.g. labour rights activists, concerned consumers, NGOs) can still feel indignation towards Atlantis and blame it.

<sup>9</sup> Being a legitimate target for reactive attitudes is sufficient for moral responsibility.

As each collective agent has its own unique ethos, at least those employees who have some leeway in how to perform their roles must have a basic grasp of corporate customs and history. Without this knowledge, they cannot ensure that they act in ways that fall within corporate culture and goals (or demand changes to it when necessary). They must also decide for themselves if the ethos is one that they are happy to work under. That said, the views of individual role-occupiers do not need to align with the collective position.<sup>10</sup> Individual role-occupiers at Atlantis need not think that health and safety are unimportant. Rather, the corporate position on minimal health and safety could have been born out of decision procedures, like where each board member gets to vote on two areas to prioritise for investment, and not enough of them prioritise health and safety. It is not that, as individuals, they do not value health and safety, it is more that as role-occupiers, such concerns are not prioritised as they try to maintain Atlantis as the go-to destination for online shoppers. The ethos of Atlantis puts a high value on customer satisfaction and prides itself on very fast deliveries. Regardless of whether the corporation could easily afford better policies, or operates on such thin margins that it can only improve one or two things at a time, Atlantis can still be blamed for prioritising the assumed needs of its customers over the wellbeing and livelihoods of its employees. Atlantis is to blame for its decision to invest only the bare minimum in health and safety.

Blame for individual role-occupiers might also make sense. Maybe the individuals had roles with enough leeway to question the collective outcome, and they could have escalated the decision higher up and or/ revised the voting rules. Alternatively, maybe the members had enough knowledge to make the case that they could have – as individuals – voted tactically to prevent the bad collective outcome. Be that as it may, I want to allow for cases where there was not enough information available for tactical voting to be a feasible option, or the members did not have enough power to be held accountable for not attempting to correct failures in the collective decision-making practices. Although individual blame would not be appropriate, blame *qua* member of a collective agent might still apply. The responses to and remedies for this are different from individual blame.

The last preliminary issue I want to note concerning collective agents is that the identity of large collective agents does not change when individuals join or leave, at least not radically or beyond

---

<sup>10</sup> Philip Pettit (2007) develops this point with his ‘discursive dilemma’ example. ‘Distributed decision making’ is proposed by Kendy Hess (2014), where individual choices by several employees result in piecemeal modifications during implementation of a corporate plan. Although each modification is small and makes sense within its limited context, put together they change the original commitment.

recognition (unless the changes in membership are sudden and large-scale).<sup>11</sup> Atlantis has had thousands of managers throughout its decade in the business. One cannot describe the identity of collective agents simply by accounting for all the individuals involved. Think of any university, especially the ones with a long history. The University of Oxford today has no overlapping members with the University of Oxford in 1478, yet it is still the same university, at least in an institutional sense. It is true that its ethos has undergone many changes over the centuries it has been in existence, such as no longer excluding women, but at the same time the history of the institution shapes the way it conducts its affairs and the kind of strengths and weaknesses it currently displays as a collective agent.

### **3. Moral actors**

The previous section explained why collective decisions and beliefs are more than just an aggregate of individual decisions and beliefs, and how corporate decision procedures, along with collective structures and policies, form the basis for corporate agency. Some philosophers link collective agency conditions strongly to moral agency: if  $x$  is a collective agent, then  $x$  is also a moral agent due to the collectivisation of reason (e.g. French 1984).<sup>12</sup> However, rational agency alone does not suffice for moral agency, even when combined with some sort of autonomy. For example, many philosophers writing on moral responsibility argue in the vein of Hume and Adam Smith that the capacity for moral sentiments is needed to feel the pull of moral reasons (e.g. Cheng-Guajardo 2019, Driver 2015, Shoemaker 2015, Strawson 1962, Wallace 1994). In this section, I will argue that collective agents do not have to qualify as moral agents in order for us to make sense of their responsibility, and suggest they should be conceptualised as moral actors instead. Moral actors maintain their own moral positions on issues and can express them. By paying rock-bottom wages and not taking health and safety seriously, Atlantis takes a moral position that the health and well-being of its employees is not a priority. Moral actors can be held fully responsible for their moral positions.

---

<sup>11</sup> A whole department resigning at once could arguably affect the identity of a collective agent, for example, but normally changes in membership happen in smaller scale and piece-by-piece. Although French et al. (1992, p. 15) argue that the corporate identity is not linked to particular people, sometimes a charismatic leader leaving can change the identity of a corporation to some degree. Additionally, the departure of a valued member of a team or a workplace bully can make a big difference to the collective ethos, but this is usually at the team or division level, rather than having an impact on the entire collective.

<sup>12</sup> The notion of moral agency of collective agents is usually supported by arguments that draw attention to the way collective decisions and actions differ from individual ones. To give an example, according to Peter French (1984), it does not make sense to put blame on individual voting members of a corporation as their views do not have to correspond with the board's decision.

The collective moral position emerges from the ethos of the collective agent and individual moral deliberations made within roles attached to the collective agent. The options and thinking of the individual members are affected by the collective in myriad ways. The members think within their roles and inside the collective framework when they help to form the collective moral position on an issue. The resulting moral outlook belongs to the collective agent: it is the collective agent's and no one else's.<sup>13</sup> The collective moral position is an emergent property of the ethos of the collective agent and the moral arguments that the (key) members make in their roles.

Essentially, I suggest that we should distinguish between the wider category of moral actors, who have the ability to maintain their own moral positions and express them through their conduct, and the more stringent moral agency conditions. Moral agency requires more than just rational agency, although rational agency is essential to distinguish between moral agents and marginal cases with moral status. After all, moral agents should be able to grasp and apply moral reasons, as well as be capable of controlling their behaviour (Wallace 1994, pp. 154-180). The former includes being able to respond to diverse situations with attentiveness, use one's own judgement, and comprehend the reasons behind justifications, while the latter means that a moral agent is able to regulate their behaviour in the light of these reasons, i.e. make choices, have some control and behave in accordance with the choices. Taking his cue from P.F. Strawson, R. Jay Wallace (1994) highlights the importance of being susceptible for moral emotions for successful exchange of moral criticism and justification. For membership in the moral community, agents must be able to grasp the *moral* reasons behind moral criticism and this goes beyond some general ability to engage in practical reasoning (Wallace 1994, p. 189). Julia Driver (2015) argues that moral agents should be moved by moral reasons and recognise moral demands, utilising a Humean account where praise and blame only have force with agents if they generate feelings of pride or humility in them. More generally, allowing for the importance of being able to engage with the attitudinal side of moral reasons, i.e. being able to feel the force of moral emotions contained in reactive attitudes, is nowadays widely accepted among moral philosophers as a feature of full moral agency (Cheng-Guajardo 2019). My argument is in line with this tradition

---

<sup>13</sup> Still, this does not necessarily let the individual off the hook, as they can have shared responsibility *qua* the collective.

and aims to show that the notion of responsibility of collective agents can accommodate also sentimental concerns over moral responsibility.<sup>14</sup>

The collective agent is able to form its own moral position through its members, who have the ability to be morally reflexive and to feel the pull of moral reasons. If none of the members of a collective agent would be moral agents, then the collective could not be a moral actor. In other words, the reflexivity of the membership is a necessary condition for moral actors and moral reflexivity is a derivative feature that collective agents as moral actors get from their members. Moral considerations of individual members feed into the collective ethos and the collective moral position. Without reflexive individual moral agency, the collective moral position would become stale, as if it were frozen in time. However, through their members who are moral agents (and as such, able to grasp and apply moral reasons to new situations), collective agents can form their own unique moral outlooks and express these. Hence, they are moral actors.

My proposal is different from that of the sceptic that Deborah Tollefsen (2003, pp. 226-228) discusses, although it might sound similar at first. She claims that according to a sceptic, we engage in dialogue with collectives only because the collective comprises of individuals with the capability for moral address. According to Tollefsen (p. 228), this amounts to failing to apprehend “the complex nature of social institutions, the ways in which authority structures and roles transform individual actions into collective actions, and the distinct properties that arise at the collective level”. The view that Tollefsen’s sceptic takes allows for a purely aggregative take of the moral outlook and positions of the collective. A reading based only on the individual members would overlook the irrefutably collective nature of the moral positions, which are not attributable to individual members. In contrast, I do not think that the *only* reason we can engage in dialogue with collectives is because they are made up of individuals with moral capacities. Like Tollefsen points out, collective positions reflect the outcome of individuals’ deliberation as role-occupiers and members of the collective. A collective moral position is never just an aggregative of individual moral positions. Therefore, collective moral responsibility is something that exists and should be discussed. Collective agents can have obligations as moral actors and we can engage in a dialogue with the collective about its outlook and moral positions.

---

<sup>14</sup> Frank Hindriks (2018) takes a different route to address sentimental concerns. He argues that the normative policies of a collective agent provides the background for members to experience collective emotions. Moral agents are conceptualised as having the capacity for moral emotions when the collectively accepted normative policies are supported by collective emotions of the members. Niels de Haan (2023) argues that this is too demanding.

But how does the ability to express moral positions through one's actions differ from simply delivering a moral message? Say that I write a message on a stone and deliver it to your door. The stone is delivering a moral message from me, but it is purely an artefact, a vessel for my message. You could say that it displays a moral position. A corporation that has accepted promoting marriage equality as a part of its ethos could display its stance, for example in a billboard advertisement, or by incorporating it as part of a storyline for a television ad. However, expressing a moral position goes further than this: the corporation is also responsive (or at least able to be responsive) to moral arguments. If I disagree with an artefact that displays a moral position or delivers a message, I get nowhere. If I disagree with an entity that has the ability to express moral positions, a moral actor, I am entering into a debate. Expressing moral positions is not just about delivering or displaying them, it is about being able to be responsive to them through the moral deliberation of individual moral agents within the collective setting.

Collective agents can express moral positions through their actions even when they are not explicitly stated. By not having an adequate health and safety policy, Atlantis is making an implicit moral claim that health and safety is not important and that it is acceptable to treat its workers this way. The collective ethos consists not only of what is deliberately revealed by the collective agent through its statements and so on, but also of what is non-deliberatively revealed through its actions. Therefore, patterns of certain practices or omissions, like repeated failures to look into an issue, can reveal a lack of regard in the collective ethos. The ethos of Atlantis includes disregard for health and safety that is manifested through the lack of action to fix issues that endanger its workers' wellbeing.<sup>15</sup>

The collective ethos is relevant for moral indictment.<sup>16</sup> Collective agents can act differently and elements of their ethos may transform. There is nothing to stop Atlantis from being more responsive to moral concerns. Although it has chosen not to put much effort into health and safety, it could change its ways. Atlantis could adopt new policies, procedures, training, incentives, and other measures to take the issue seriously from now on, perhaps in response to moral criticism. Without such changes to the ethos being possible, it would not make sense to praise or blame a collective agent, as it would mechanically continue on its path regardless of the response it is getting. Maybe some managers or board members within Atlantis are trying to push the corporation to take employee safety more seriously, it is just that there are not enough

---

<sup>15</sup> It could also be due to bad practises with no ill will, but then the collective agent should act once they are alerted to the problem. If not, disregard for such matters is part of their ethos.

<sup>16</sup> The collective ethos can condition members in both good and bad ways.

of them to create change at the collective level. So far, this increasing divergence of opinion among managers or board members is not reflected in the decisions taken by Atlantis.

Collective agents can also claim to have a certain ethos, but be misleading about it, whether deliberately or not. When a collective agent expresses moral positions through their conduct, their sincerity can be assessed, just like an individual moral agent's. What I have in mind is tantamount to a dishonest individual. This dishonest agent might, for example, publicly claim to be for gender equality in the workplace but is then caught saying or doing misogynous things. In the same way, a company could publicly advertise its commitment to respecting workers' rights, only to be found out to be the largest client of a decrepit factory that is at risk of collapsing on its workers. The fact that the company has made the purchasing orders through its subsidiary does not make the claim of respecting workers' rights any less dishonest (perhaps it has even tried to actively make its operations non-transparent by using subsidiaries).<sup>17</sup> In cases like these, we can (and do) state that the company does not value the health and safety of the workers in its supply chain highly enough. Valuing such things does not appear to be part of its ethos. Through their actions, collective agents express and reveal the values that really constitute their ethos. Although Chez Louis prides itself on sourcing ethical baking supplies, the true scope of social responsibility within its ethos is only revealed after we know how it treats its own workers, for example. Sometimes discrepancies between proclaimed values and actions are not due to dishonesty; sometimes it is a matter of moral confusion or not thinking through one's values. In a similar way, a collective actor can express conflicting values, or their actions can betray their ethos as they have not considered all of the consequences properly. In these cases, their response to blame reveals the strength of the values that they actually decide to promote.

Moral actors could also be described as thin moral agents as they lack the thicker qualities (reflexivity, moral sentiments) that constitute fully fledged moral agency. However, I prefer to distinguish these collective creatures with the new label of a moral actor as any "thinness" in their moral agency is not due to them somehow being incapable of making sound moral judgements and justifiable moral claims. They are capable of this through their members: the moral outlook and positions of collective agents as moral actors are an emergent property. A moral actor can maintain and express moral positions that are implicit or explicit in their ethos. The ethos includes both the stated and unstated values of the collective agent. The actions and

---

<sup>17</sup> The moral positions implicit or explicit in collectives' ethos can manifest through the decisions they make regarding their supply chains and subsidiaries.

inactions of the collective agent reflect and reveal its collective ethos, and it is the ethos that the protesting workers are responding to.

My argument in this section has been that we should regard collective agents as moral actors, part of our moral sphere. Although a collective agent's moral reflexivity derives from its members' capacity for moral reflexivity, there is nothing second-order about claims of collective moral responsibility. Collective agents can maintain and express their own moral positions and be responsive to moral concerns. We should not demand less morally from collective agents than from other agents. They can be held fully morally responsible.<sup>18</sup> This is an important consequence of my argument.

#### **4. A functionalist alternative and its limitations**

An alternative route to avoid the problematic notion of corporate moral agency without moral emotions is to conceptualise these in purely functionalist terms. Gunnar Björnsson and Kendy Hess (2017, p. 274) have argued that if a collective is “capable of agency, then they are also capable of *states sufficiently similar to guilt and indignation to satisfy the requirements of moral agency*”. Essentially, they suggest that corporate positions are equivalent to beliefs, corporate goals function like desires, and corporate plans are functionally equivalent to intentions. What matters, according to Björnsson and Hess (2017, p. 288), is whether corporate agents “can have the properties required for fully fledged moral agency”, not if they can closely resemble human agents.<sup>19</sup> I have nothing against discussing corporate “mental states” in functionalist terms. However, I think the functionalist account of reactive attitudes of collective agents is too thin to account for cases like the ones I have been discussing. This is because the proposal of the moral equivalency of reactive attitudes relies on an unrealistically rational and mechanistic view of corporate agency. I will elaborate in this section on why I think this is so. In my criticism, I hope to do justice to the careful and detailed way that Björnsson and Hess proceed with their interesting argument.

---

<sup>18</sup> Depending on circumstances and context, mitigating factors might of course apply. Still, the point remains: unlike marginal moral agents, there is nothing to prevent holding collective actors fully responsible.

<sup>19</sup> Björnsson and Hess (2017, p. 288) acknowledge that completely developed moral agency could entail “the phenomenology of reactive attitudes”.

The authors grant that the corporation's individual members might well hold incoherent commitments and that there are disagreements about these among the members. Still, when the individuals act as members, namely within their roles, they are guided by the corporate commitments to act in reliable ways. According to their model, a corporation's positions, goals and plans (i.e. beliefs, desires and intentions) "do not blatantly contradict each other", but instead "form a rational, coherent profile" and "a logically integrated complex of commitments about fact and value that drive corporate action" (p. 278). This is obviously intended as an idealised version of how a collective agent works, to abstract away from the times when the corporate machine is not operating so smoothly in relation to planning and goal-setting (i.e. almost always). In the idealised model, subordinates prepare proposals for the higher-ranking members, which culminates in the board of the corporation voting for a proposal: the one with the majority of votes wins and is incorporated as a new corporate position. Still, even as an idealised version, this mechanistic view of collective decision-making operates on too narrow a terrain to really appreciate the messiness and the inherent power struggles involved in it.

In reality, the corporate position on an issue is a richer entity of what is adopted in the written format, be that meeting minutes, guidelines or some other document, or even what are generally included as the company's norms. The individual desires, beliefs, commitments and ideals of its members influence the ethos of a collective in many ways. A strong ideological push from a handful of members can sometimes result in a collective agent adopting a position on an issue that does not seem rational, or at least is far from an obvious choice. In addition, issues like the company culture or the wider norms in society affect the kind of decisions that individual role-occupiers push for. Nor does planning need to proceed in an orderly, rational fashion. Again, some members might, either covertly or openly, push for things to be included in the plan that others might not agree with, regardless of their merit. Higher-level role-occupiers can respond in different ways to the decisions and interpretations made by the subordinates in response to the collective upper-level plan, or the *masterplan* as I will call it here.

Think of a corporate commitment to implement equal opportunities across all 200 warehouses owned and operated by Atlantis. The masterplan to do this can be detailed or simple, but the written plan by itself is not yet enough to conclude in one way or another whether the corporation's ethos now truly embraces equal opportunities. The executive level and managers at Atlantis can adopt a very relaxed attitude towards the various microlevel plans that flow from the macrolevel masterplan. This translates into the collective ethos being lax when it comes to

implementing and supervising equal opportunities. One could then justifiably blame Atlantis for only paying lip service to equal opportunities. There are countless real-life examples to be found of such a mismatch between stated or advertised corporate goals and what really takes place.

Although Björnsson and Hess seem to have inherent optimism about corporate agents acting rationally, in a footnote they admit that this need not be the case, and that positions, goals and plans can also be formed in a less rational and more haphazard way.<sup>20</sup> Therefore they might not disagree with what I have just said. Even so, they seem to think that the rationalistic model is the preferred model for how things should proceed. In contrast, I think that collective decision-making that adheres to rational and procedural norms is only half the story and is not necessarily in any way ideal. Just think of a rationally and procedurally well-functioning bureaucracy that serves morally bad ends in a dictatorship.

To be clear, rationality and procedures clearly play a major role in planning and I grant that in many cases a rationally coherent relationship between subplans and masterplans is desirable. As Shapiro (2011) has argued, there would be no point in planning if we kept dropping our previous (planned) commitments and kept planning afresh each day. This would be wasteful, as well as make any collective action or cooperation difficult, if not impossible. Still, rationality and procedures will only take us so far, even in collective settings. What is needed in addition is moral reflexivity, and this comes from outside the planning. Naturally places for moral reflexivity can be noted in a plan (“review regularly that equal opportunities are implemented across all branches”), but the actual reflexivity is never mechanical. It cannot be detailed in advance, even in the best policies and procedures. Think of cases where a new form of bullying emerges and needs to be recognised as such by the moral community (e.g. new forms of cyberbullying that have sprung up during the last decade). A moral agent needs to respond to novel cases with the understanding and empathy of a fellow moral agent.

Moreover, allowing for a messier and less rational way for corporate moral commitments to form seems to pose a problem for the Björnsson and Hess account. Discussing an example about a corporation that has unknowingly polluted a river next to their factory, they argue that

---

<sup>20</sup> In footnote 16, Björnsson and Hess (2017) argue that while corporate positions and plans might be shaped by sound deliberation and information-gathering, they could also “result from miscommunication or from individual members abusing the system by pushing their own interests in violation of formal requirements”. They also concede that the commitments might even “arise organically out of shifts in member behavior that are not even aimed at shaping larger corporate behavior or commitments”. However, in the main text they do not explore the implications of this less rational way of arriving at corporate commitments.

the way the corporation springs into action as a result can be explained by it having a position that it is guilty of the pollution. My worry is, in a nutshell, that this explanation seems to rely heavily on the mechanistic model of corporate decision-making that I have argued is too simplistic, at least for the kind of cases of corporate blame that I seek to discuss. Essentially, in their example, the corporation recognises its responsibility in a “polluter-pays” kind of way: its actions polluted the river, so it should clean up the mess. Questions of blame and guilt are irrelevant for admission of a responsibility of this sort. To have an apology mechanistically triggered by a recognition of harm is akin to a robot dealing with workplace bullying. This is why I think that in their example, compensation and the issuance of an apology are red herrings. If compensation and apologies follow mechanistically from a previous strategic decision to have a green company image, and no member feels any guilt in their role on behalf of the company, there is no real apology and no real making of amends taking place. Rather, the responsibility accepted is purely causal. It is then a question relating solely to the existing policies and procedures as to whether the polluting corporation pays of its own volition or only through a legal process.

In contrast, when protesters blame Atlantis for mistreating its workers, or when students are saying that their school is not taking sexual harassment claims seriously enough, what they are essentially trying to achieve is to get the collective agent to change its ethos in relation to their demands. The kind of change they are after is about expanding the realm of concern of the collective agent, or about challenging the answers it has previously accepted as part of its intentional horizon. Mechanistically responding along procedural lines will not suffice when it comes to responding to such claims.

Furthermore, many interesting cases of blame directed towards a collective agent are ones where a new moral situation arises, be it from changes in our social realities or practices, a new kind of harm in search of a name, societal or institutional norms being pushed, or pain stemming from old structural harms that are still not being properly addressed, to mention just a few examples. In cases like these, existing procedures and policies will not suffice, as nothing will change if people keep following the status quo. The changes demanded by the blame are also seldom easy to implement in a quick-fix way. There is often no blueprint to follow about what should be done. What is required instead is open moral deliberation. This kind of deliberation takes time and resources. It is hard to see how a collective agent would proactively dedicate the said time and resources to think through its position towards an issue where the members feel no moral

pull from guilt, for example. Acknowledgement of guilt, or the threat of other agents judging you blameworthy, can be a powerful motivator. Existing corporate mechanisms might be able to process polluter-pays cases, but I cannot see how they could succeed in dealing with the more complex cases requiring moral introspection.

The wider point is that while the functionalist model might be enough to deal with cases where the path for responsibility is clearly laid out, it has no bite with more complex demands and reactive attitudes towards collective agents. I believe that the account I have presented in this chapter puts corporations on the hook just as much as the functionalist account does, but that it can also deal with the messier reality of blaming corporations. It is also simpler in its structure and comes closer to our common-sense morality. What reactive attitudes tell us about our practice of holding agents responsible is that we expect moral agents to not only rationally understand praise or blame as judgements, but also to feel their motivational pull. As we have other ways of appreciating the collective constraints and context that the individual members' moral agency operates under, we do not have to muddy the waters of moral agency by conceptualising collectives as such.

## **5. Concluding remarks**

My argument has been that collective agents, such as corporations and universities, act with moral valence and should be understood as moral actors in our social world, able to maintain and express their own collective moral positions. As such, they can be held morally responsible for their acts and omissions. Conceptualising collective agents as moral actors can account for concerns of moral sentimentalists, as well as aligns our theories closer to our common-sense morality. This could make it harder to distance oneself psychologically from the effects of one's acts carried out within a collective setting. More importantly, it allows us to discuss the responsibility of collective agents without treating them as moral patients or marginal agents that have only some degree of responsibility. How well collective agents act morally is up to their structure and ethos. These are not set in stone, which could potentially motivate some key individuals in collective agents to act differently within their roles to try to push for a change for the better. It also provides a simple theoretical background for those who want to blame or praise actions by collective agents.<sup>21</sup>

---

<sup>21</sup> This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No GAP-839448. I wish to thank the anonymous referee and my co-

## Bibliography

- Arpaly, N. (2015) *Huckleberry Finn* Revisited: Inverse Akrasia and Moral Ignorance. In R. Clarke, M. McKenna & A.M. Smith (Eds.), *The Nature of Moral Responsibility: New Essays* (pp. 141–156). New York: Oxford University Press.
- Bakan, J. (2004) *The Corporation: The pathological pursuit of profit and power*. New York: Free Press.
- Björnsson, G. & Hess, K. (2017) Corporate Crocodile Tears?: On the Reactive Attitudes of Corporations, *Philosophy and Phenomenological Research*, 94, 273-298.
- Cheng-Guajardo, L. (2019) Responsibility Unincorporated: Corporate Agency and Moral Responsibility, *The Philosophical Quarterly*, 69, 294–314.
- Collins, S. (2022) I, VOLKSWAGEN, *The Philosophical Quarterly*, 72, 283–304.
- de Haan, N. (2023) Collective moral agency and self-induced moral incapacity, *Philosophical Explorations*, 26, 1-22.
- Driver, J. (2015) Appraisability, Attributability, and Moral Agency. In R. Clarke, M. McKenna & A.M. Smith (Eds.), *The Nature of Moral Responsibility: New Essays* (pp. 157–173). New York: Oxford University Press.
- Erskine, T. (2003) Introduction: Making Sense of ‘Responsibility’ in International Relations – Key Questions and Concepts. In T. Erskine (Ed.), *Can Institutions Have Responsibilities? Collective Moral Agency and International Relations* (pp. 1-16). Basingstoke: Palgrave Macmillan.
- French, P. A. (1984) *Collective and Corporate Responsibility*. New York: Columbia University Press.
- French, P. A., Nesteruk, J. & Risser, D. T. with Abbarno, J. (1992) *Corporations in the Moral Community*. Fort Worth: Harcourt Brace Jovanovich College Publishers.
- Hess, K. M. (2014) Because They Can: The Basis for the Moral Obligations of (Certain) Collectives, *Midwest Studies In Philosophy*, 38, 203-221.
- Hieronymi, P. (2001) Articulating an Uncompromising Forgiveness, *Philosophy and Phenomenological Research*, 62, 529-555.
- Hindriks, F. (2018) Collective Agency: Moral and Amoral, *Dialectica*, 72, 3-23.
- Isaacs, T. (2014) Collective Responsibility and Collective Obligation, *Midwest Studies In Philosophy*, 38, 40–57.
- Kennett, J. (2006) Do Psychopaths Really Threaten Moral Rationalism?, *Philosophical Explorations*, 9, 69–82.
- Laitinen, A. (2014) Collective Intentionality and Recognition from Others. In A. Konzelmann Ziv & H. B. Schmid (Eds.), *Institutions, Emotions, and Group Agents: Contributions to Social Ontology* (pp. 213–227). Dordrecht: Springer Science+Business Media.
- Pettit, P. (2007) Responsibility Incorporated, *Ethics*, 117, 171–201.
- Riskin, J. (2016) *The Restless Clock: A History of the Centuries-Long Argument over What Makes Living Things Tick*. Chicago: University of Chicago Press.
- Shapiro, S. J. (2011) *Legality*, Cambridge, MA: Harvard University Press.
- Shoemaker, D. (2015) *Responsibility from the Margins*. Oxford: Oxford University Press.

---

editor Bill Wringe for their useful feedback and criticism. I am also grateful to Lilian O’Brien and Christopher Kutz for their astute comments on an earlier version. Holly M. Smith and R. Jay Wallace offered helpful suggestions during a virtual meeting in March 2020, conducted to replace my *Berkeley Law’s Visiting Scholars Program Workshop* presentation (cancelled due to the pandemic). I would also like to thank the audiences at *MANCEPT 2021* workshop *The Normative Profile of Collective Agents* and *APA Eastern Division 2022* symposium session, particularly my commentator Alex R. Steers-McCrum. I also received perceptive comments in the *Empirical approaches to rationality and action group* in 2021, with special thanks to Polaris Koi for written comments. In addition, I would like to thank the audience members at the *Collective Responsibility workshop* at the University of Helsinki in November 2017 where I presented an early draft of the paper, especially Gunnar Björnsson and Stephanie Collins. Last but not least, I thank Marion Godman, Arto Laitinen, Holly Lawford-Smith and Pekka Mäkelä for providing me with written comments on the first draft of the central idea, written back when I was on a personal research grant from the Kone Foundation.

- Smiley, M. (2022) Collective Responsibility. In E. N. Zalta & U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy (Winter 2022 Edition)*, plato.stanford.edu/archives/win2022/entries/collective-responsibility.
- Smith, A. M. (2013) Moral Blame and Moral Protest. In D. J. Coates & N. A. Tognazzini (Eds.), *Blame: Its Nature and Norms* (pp. 27–48). New York: Oxford University Press.
- Srinivasan, A. (2018) The Aptness of Anger, *Journal of Political Philosophy*, 26, 123–144.
- Strawson, P. F. (1962) Freedom and Resentment, *Proceedings of the British Academy*, 48, 1–25.
- Tollefsen, D. P. (2003) Participant Reactive Attitudes and Collective Responsibility, *Philosophical Explorations*, 6, 218–234.
- Tuomela, R. (2007) *The Philosophy of Sociality: The Shared Point of View*. New York: Oxford University Press.
- Tuomela, R. & Mäkelä, P. (2016) Group Agents and Their Responsibility, *The Journal of Ethics*, 20, 299–316.
- Véliz, C. (2021) Moral zombies: why algorithms are not moral agents, *AI & Society* 36, 487–497.
- Wallace, R. J. (1994) *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.