The Philosophical Quarterly Vol. 0, No. 0 *2023*
ISSN 0031-8094 https://doi.org/10.1093/pq/pqad104

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

GROUP RESPONSIBILITY AND HISTORICISM

By Stephanie Collins[1] and Niels de Haan[2]

In this paper, we focus on the moral responsibility of organized groups in light of historicism. Historicism is the view that any morally responsible agent must satisfy certain historical conditions, such as not having been manipulated. We set out four examples involving morally responsible organized groups that pose problems for existing accounts of historicism. We then pose a trilemma: one can reject group responsibility, reject historicism, or revise historicism. We pursue the third option. We formulate a Manipulation Condition and a Guarding Condition as addendums to historicism that are necessary to accommodate our cases of group responsibility.

Keywords: historicism, corporate responsibility, group agency, group responsibility, organized groups, moral responsibility, structuralism.

I. INTRODUCTION

We regularly hold organized groups, such as corporations or states, morally responsible. In recent decades, philosophers have sought to vindicate this practice. These philosophers have argued that some groups can robustly form and act upon representational and goal-seeking states. These groups are agents. Further, (a subset of) these groups can understand moral reasons and act accordingly. The idea is that group-level responsibility is non-redundant, and neither entails nor precludes the responsibility of members. This idea has achieved a high level of consensus (French 1984; Isaacs 2011; List and Pettit 2011; Bjornsson and Hess 2016; Hindriks 2018).

Yet organizations are designed and influenced by other agents. Organizations are designed and influenced to have specific decision-making mechanisms, organizational structures, and constitutive ends—which affect their decision-making and action-taking. A seemingly responsible group cannot always change these imposed mechanisms, structures, and ends. This poses problems for anyone who wants to endorse both of two popular views: first, the view that organized groups are morally responsible (in certain cases), and

© The Author(s) 2023. Published by Oxford University Press on behalf of The Scots Philosophical Association and the University of St Andrews. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

second, 'historicism' about responsibility. Historicism says that any responsible entity must satisfy historical conditions, such as lacking a history of relevant manipulation. Like groups' responsibility, historicism is popular amongst contemporary philosophers (Fischer and Ravizza 1998; Haji and Cuypers 2007; Haji 2013; McKenna 2016; Mele 2019).

We unite these two disconnected literatures, arguing that certain cases of organized groups' responsibility undermine existing historicist principles. This raises a trilemma: we can abandon attractive judgements about group responsibility, abandon historicism, or revise historicism. Our main aim is to raise this trilemma as an important choice point in the philosophy of responsibility. That said, we favour the third option. We therefore sketch two revised historicist principles, which capture the judgements concerning group responsibility.

Our argument proceeds as follows. In Section II, we outline three existing historicist principles. In Section III, we provide four examples of group responsibility, which challenge these three historicist principles. In Section IV, we present the trilemma: reject our examples, reject historicism, or revise historicism. We advocate the third option, sketching two new historicist principles. This brings the two literatures into harmony.

Before beginning, a word on methodology. We assume principles of responsibility should cover all agents—individual and collective. We privilege neither individual nor group agents in our reflective equilibrium. Instead, we pursue a reflective equilibrium that accommodates judgements about both agent-types, with principles that are neutral between individuals and groups. We therefore reject, with Sara Rachel Chant (2021), the 'wash, rinse, repeat' approach to group agency, in which theories of agency and responsibility start from individuals as the paradigm case, and mechanically apply the resulting theories to collectives. We assess existing historicist principles by their applicability to group agents, even if their authors would reject such an application (e.g. Haji (2006) and McKenna (2006)).

II. HISTORICISM

Historicism's rival is structuralism. Under structuralism, whether an agent is morally responsible for an action depends only on her psychological structure when she performs the action (Wolf 1987; Frankfurt 1988; Watson 2004; Vargas 2006; Cyr 2020). Under historicism, whether an agent is morally responsible for an action partly depends on her history before the action (Fischer and Ravizza 1998; Haji and Cuypers 2007; Haji 2013; McKenna 2016; Mele 2019; technically, the debate concerns the 'agency' condition on responsibility. Responsible agents must also satisfy the 'epistemic' and 'control' conditions).

Both camps agree that history can matter for 'indirect' responsibility. For example, suppose you freely decide to get drunk to test your drunk-driving

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

abilities. Your responsibility for hitting someone while drunk is 'indirect': it traces to your free choice to test your drunk-driving, even if you do not meet conditions for responsibility when you hit the person. Structuralists and historicists agree that your *direct* responsibility for past choices can affect your *indirect* responsibility for present outcomes. What they disagree about is the relevance of history for *direct* responsibility, such as your responsibility for choosing to get drunk and test your driving.

Structuralists say an agent's direct responsibility is solely determined by how she was structured at the time, regardless of how she came to be structured that way. For example, according to Harry Frankfurt, you are responsible for your choice to get drunk if you 'identify' with the choice when you make it. You are responsible for your choice even if your identification derived from manipulation, socialization, indoctrination, and so on (Frankfurt 1988: 171–2). We will discuss advantages of structuralism in Section IV.2.

On historicism, how an agent became how she is at a time bears on whether she is responsible for her actions at that time (Levy and McKenna 2009). The debate centres on *manipulation cases*. Two prominent such cases motivate the three existing historicist principles we outline below. Mele (2016) calls these 'radical reversal' and 'original design' cases, respectively. First, the radical reversal case:

> **Beth.** Ann is a free agent and an exceptionally industrious philosopher. […] Beth, an equally talented colleague, values many things above philosophy for reasons that she has refined and endorsed on the basis of careful critical reflection over many years. […] Their dean wants Beth to be like Ann. […] Without the knowledge of either philosopher, he hires a team of psychologists […and] new-wave brainwashers […] The psychologists decide that Ann's peculiar hierarchy of values accounts for her productivity, and the brain-washers instill the same hierarchy in Beth while eradicating all competing values … Beth is now, in the relevant respect, a 'psychological twin' of Ann […] Ann, by hypothesis, freely does her philosophical work, but what about Beth? (Mele 2006: 164–6)

Historicists hold that Beth is not acting freely and is not morally responsible for any subsequent wrongdoing (even if she identifies with the wrongdoing at the time) because she is brainwashed.

Second, the 'original design' case:

> **Suzie.** Suzie is created by a god at an instant. Suzie is created to be a psychologically healthy woman indistinguishable from any other normal functioning thirty-year old. She is given a huge set of beliefs according to which she has lived a normal human life for thirty years. Furthermore, Suzie has some range of values and principles that are unsheddable.[1] She has a false set of beliefs about how she came to acquire her

[1] We assume 'unsheddable' values cannot be removed by will. An agent with unsheddable values can be self-controlled, reason-responsive, and override their unsheddable values on occasion (supposing this does not involve losing the values entirely).

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

> unsheddable values. She is a richly self-controlled person who is able to resist the inclination to act with weakness of will. She is reason-responsive (sans any historical component). Suppose Suzie is presented with the option to do A or B. Option B involves a violation of a value that is unsheddable for her. Option A involves acting from one of her unsheddable values. Suzie A-s, acting as her unsheddable value counsels, but in doing so, she could have done otherwise—that is, she could have B-ed. (McKenna 2004: 180–1)

McKenna thinks it natural to say Suzie A-ed freely. After all, he says, if the same god created a different person at birth, who grew up to be a duplicate of Suzie, that person would be responsible. Thus, the god's 'original design' does not undermine Suzie's responsibility.

Obviously, Beth and Suzie are highly artificial and outlandish examples. Without endorsing the use of such examples, or the epistemic status of judgements about them, we use these examples to illustrate existing historicist principles. Historicists have tried to develop principles that give explanatory verdicts on these examples. We discuss three such principles.

First, Mele proposes the following sufficient condition for an agent's non-responsibility for an action:

> **Mele's Principle.** An agent does not freely A and is not morally responsible for A-ing if the following is true: (1) for years and until manipulators got their hands on him, his system of values was such as to preclude his acquiring even a desire to perform an action of type A, much less an intention to perform an action of that type; (2) he was morally responsible for having a long-standing system of values with that property; (3) by means of very recent manipulation to which he did not consent and for which he is not morally responsible, his system of values was suddenly and radically transformed in such a way as to render A-ing attractive to him during t; and (4) the transformation ensures either (a) that although he is able during t intentionally to do otherwise than A during t, the only values that contribute to that ability are products of the very recent manipulation and are radically unlike any of his erased values (in content or in strength) or (b) that, owing to his new values, he has at least a Luther-style inability during t intentionally to do otherwise than A during t.[2] (Mele 2019: 66–7)

On Mele's Principle, responsible agents must *lack* a certain history: he endorses 'negative' historicism. Beth has the history Mele describes, so is not responsible on Mele's view. Suzie hasn't existed for years, so lacks the history Mele describes. Mele's Principle is therefore silent on Suzie's responsibility. Mele can presume Suzie is responsible, pending further sufficient conditions for lacking responsibility.

Haji and Cuypers (2007; similarly Haji 2013) also defend negative historicism. They focus on 'evaluative schemes,' consisting of (1) normative

[2] We interpret a 'Luther-style' inability as the inability expressed by Martin Luther's apocryphal statement 'here I stand, I can do no other.' It's the inability felt by someone with strong values, though they can physically do otherwise.

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

standards the agent believes ought to be invoked in assessing reasons for action or beliefs about choices; (2) desires, beliefs, or plans expressing the agent's long-term ends or goals he deems worthwhile or valuable; (3) deliberative principles the agent uses to arrive at practical judgements; and (4) motivation to act on the normative standards specified in (1) and pursue goals described in (2) at least partly based on the principles in (3) (Haji and Cuypers 2007: 350). They endorse:

> **Haji and Cuypers' Principle:** An agent A is an appropriate candidate for moral responsibility for action only if, at the time A performs the action, A has (i) an evaluative scheme that is *responsibility-wise authentic* (at least with respect to the constituents from which the action is issued); (ii) deliberative skills and capacities; and (iii) executive capacities. (Haji and Cuypers 2007, not a direct quote)

There are two kinds of 'responsibility-wise authentic evaluative schemes': an *initial* scheme (the first scheme an agent has; consider Suzie) and an *evolved* scheme (a scheme held later in life; consider Ann and Beth). An *initial* scheme is 'responsibility-wise authentic' if it does not compromise the agent's being morally responsible at future times (2007: 370). An *evolved* scheme is 'responsibility-wise authentic' if it resulted from acceptable modifications—made under A's deliberative control—to doxastic and motivational constituents of a prior evaluative scheme that was responsibility-wise authentic.

Haji and Cuypers do not require a responsible agent to have a past. But if one has an evolved evaluative scheme, it must *lack* a history of modifications that bypassed the agent's deliberative control. Like Mele's Principle, Haji and Cuypers' Principle renders Beth non-responsible. Haji and Cuypers note that Suzie's responsibility depends 'on how the tale is spun'. They 'assume' Suzie's initial evaluative scheme relates appropriately to her later evaluative scheme, facilitating her responsibility for initial actions (2007: 359).

Under positive historicism, responsible agents must *have* a responsibility-enabling history, rather than *lack* a responsibility-disabling history. Consider:

> **McKenna's Principle**: An agent performs a directly free act and is directly morally responsible for it only if any unsheddable values playing a role in the production of her action arose from a history whereby she was afforded the opportunity to critically assess, endorse, and sustain them from abilities that she possessed, and so none were acquired through means that bypassed those abilities. (McKenna 2016: 97)

McKenna's Principle renders Beth non-responsible, but also renders Suzie non-responsible: Suzie lacks a history, including a history of opportunities for critical reflection, so she doesn't satisfy McKenna's necessary condition on responsibility. In defence of this, McKenna notes that Suzie's 'evaluative framework was fully fixed for her without her so much as having a chance to have engaged in the shaping of it' (2016: 98). We'll show that such a denial is more difficult in Suzie-style cases involving organized groups.

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

There is another influential historicist theory: that of Fischer and Ravizza (1998). Their theory is 'subjectivist': it requires a responsible agent to have a history of *viewing itself* as an agent and an apt target of reactive attitudes. Subjectivist historicism cannot attribute responsibility to organizations that never view themselves in this way. Yet many organizations face economic or social incentives never to take moral responsibility. Indeed, these organizations commit the worst offences: it is implausible that they are never morally responsible. Fischer and Ravizza do discuss agents who deny their own responsibility (1998: 217–20), arguing that the price of avoiding responsibility is too high for humans to pay: it causes 'sequestration' from important human relationships, with disastrous psychological consequences. However, many organizations couldn't care less about important human relationships or sequestration from the relational community. Therefore, we put aside Fischer and Ravizza's theory. To capture all organizational agents, we need a non-subjectivist version of historicism. Again, here we assume that historicist principles must cover all agent-types: perhaps Fischer and Ravizza's theory applies to humans, but it cannot apply to organizations.

With these three historicist principles on the table, we turn to organized groups.

III. ORGANIZED GROUPS AND GROUP RESPONSIBILITY

We assume the groups we'll discuss qualify as agents capable of understanding and processing moral reasons and acting accordingly. Can historicism accommodate organizations' responsibility?

There are (at least) three features of organizations that are liable to Beth-style manipulation or Suzie-style design: (1) decision-making procedures, (2) organizational structures, and (3) constitutive ends. A group can be responsible for actions resulting from these features. This is the core insight of philosophical research into group agency and responsibility in recent decades.

For example, Christian List and Philip Pettit (2011) focus on *decision-making procedures*. They explain how a group's decision-making procedure might lead the group to an immoral decision, despite no members being such that they would make that decision if they alone were deciding for the group. List and Pettit argue the group is responsible for the decision. Peter French (1984) focuses on *organizational structures*. He explains how inadequate structures of reporting, oversight, and command can lead to immoral group actions, for which the group is responsible. Finally, Carol Rovane (1998) focuses on (what we call) *constitutive ends*. She argues that groups have a rational point of view unified by a central life project, which might be discontinuous with the projects of members. The group is responsible for the pursuit of its project.

Following these authors, we focus on groups with high levels of organization, structure, and integration. We leave aside looser groupings, which are

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

much more contentious as bearers of 'agency' and 'responsibility' (e.g., the loose social groupings theorized by Gilbert 2014 or Graham 2002). In many situations, organized groups can conform to the historicist principles outlined above. Those principles require the organization to have had (or at least not have been denied) the opportunity to *reflect* on its decision-making procedures, organizational structures, and constitutive ends. In many situations, organizations are responsible, under the above historicist principles.

However, there are examples where organizations are *not* responsible under the above principles, yet where philosophers may want a theory that explains how the organization *is* responsible. Some cases concern a group's decision-making procedure and organizational structure: its *formal features*. Other cases concern a group's constitutive ends: its *substantive features*. Using this formal/substantive distinction, we will present four cases: a radical reversal case involving formal features; an original design case involving formal features; a radical reversal case involving substantive features; and an original design case involving substantive features.

Our 'formal' cases demonstrate that an agent's responsibility is influenced not only by the agent's *values* (and how the values came about), but likewise by the *mechanisms* the agent uses to make decisions (and how those mechanisms came about). Decision-making procedures and structures are crucial parts of (group) agency. These features can be subject to manipulation in ways that matter for historicism, just as values can. One key insight of our examples is that the importance of procedures and structures has been overlooked by the focus on *values* within the historicist literature. To be sure, the manipulation of procedures and structures can affect an entity's values 'downstream'. But such downstream effects are not the only reason why manipulation of procedures and structures is important: such manipulation can undermine responsibility, even if it doesn't affect values.

Another word on our examples. Our examples are more technologically plausible—and more embedded in existing practices—than the outlandish examples of Beth and Suzie. Indeed, our examples are inspired by real-world cases in which organizations were held responsible. Thus, it is plausible that our best theories of responsibility should produce responsibility in the upcoming examples—even if our best theories might prevaricate on the responsibility of (say) Suzie. As with all philosophical thought experiments, our examples exclude many potentially relevant details, including the full life histories and complex social contexts of the individuals involved. Our point is that these details *could* be filled in, such that historicists and other philosophers would want to hold the organization responsible.

First, consider the following 'radical reversal' case:

> **Safety**: At t_1, and for many years prior, an organization prioritizes worker safety. The organization arrived at that value via long-term rational reflection on competing

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

> values. At t_2, the organization is sent into involuntary administration, through no fault of the organization. The administrator imposes a new decision-making procedure on the committee that oversees production: the committee will now vote only on premises, where those premises entail certain conclusions. The administrator has seen this procedure work well in many other contexts. The committee votes on three premises: that safety-engineer hours will be decreased; that machine maintenance will be delayed; and that factory output will be increased. Any two of these premises would not be sufficient to compromise worker safety. But together, these three premises do compromise worker safety. Each premise gets majority support. Via its new premise-based decision-making procedure, the committee decides to compromise worker safety. However, no member supports that conclusion. Each member voted for only two of the premises, though the votes were spread such that all three premises got majority support. At t_4, the committee's decision leads to an accident in which several workers die.[3]

The organization is not responsible, by existing historicist principles. It was subject to a recent manipulation of its decision-making procedure, to which it did not consent. This manipulation *indirectly* affected the organization's values: radical reversal cases are as much about manipulating *how* one makes decisions, as manipulating *what* one values. The organization meets (a refinement of) Mele's sufficient condition for non-responsibility (where the refinement accommodates manipulation of formal features): the organization could decide otherwise, but decides this way because of a manipulation that altered its decision-making process. Regarding Haji and Cuypers' Principle, the administrator's interference influenced the organization's 'evolved evaluative scheme', while bypassing the organization's capacities of deliberative control. Likewise for McKenna's Principle: the history from which the organization's later (safety-compromising) values arose did not afford the organization the opportunity to critically assess, endorse, and sustain those values.

Historicists might accept the organization's non-responsibility. Perhaps responsibility lies, instead, *solely* with the administrator or the previous directors. We examine this possibility in Section IV.1. But, prima facie, there is strong presumptive reason to judge that the organization is responsible in Safety. After all, the organization's decision led to the deaths of several workers. This is clearly morally wrong. The organization could have ensured the safety of its workers, and it surely should have known that decreasing safety-engineer hours, delaying machine maintenance, and increasing factory output would endanger its employees. The only question is whether the manipulation is so complete that the 'agency' condition on responsibility is not satisfied. In

[3] The vote creates a 'discursive dilemma,' for which List and Pettit (2011) argue organizations are responsible. Peter French (1984) analyses the real-world example of Air New Zealand's crash on Mount Erebus, in which the organization's decision-making procedure caused 257 deaths. Mount Erebus did not involve manipulation or voting mechanisms, but it demonstrates how the good-faith use of poor organizational procedures can create moral calamity and generate group responsibility.

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

Section IV.3, we argue that the manipulation is insufficient to exculpate the organization; therefore, the agency condition is satisfied in this case (unlike in the case of Beth).

Our second case involves original design.

> **Disaster:** At t_1, a mining corporation is created with three departments: Dig, Extraction and Supply. To expediate decision-making, each department is given a budget and broad autonomy. At t_2, the corporation decides to build a large mine close to a dam. Unfortunately, the corporation's communicative structure functions too slowly to affect the real-time decision-making of the relevant departments, who operate under time pressure and must meet deadlines. Each department cuts costs that lead to the erosion of earth. Each department's contribution is neither necessary nor sufficient for causing harm. Together they cause the earth to erode such that the dam collapses. The collapse results in dozens of deaths of locals and a huge environmental disaster. The disaster was avoidable within the budget constraints.[4]

Disaster is a counterexample to historicist principles that *deny* the responsibility of newly created agents, like Suzie. Consider McKenna's Principle. The corporation lacked the opportunity to critically assess its communicative structure, since this was the first time it was employed. The corporation was not afforded an opportunity to critically assess this feature that significantly impacted its agency. It is therefore non-responsible, under McKenna's Principle. Again, Section IV.1 will consider accepting this result. For now, we assume some historicists will want to assert responsibility in *Disaster*. Again, there is clear organizational wrongdoing, the disaster was avoidable, and the organization should have known the severe erosion could lead to the dam's collapse. In Section IV.3, we will explain how Disaster differs from Beth.

Our third example concerns *constitutive ends*. These relate to an agent's practical identity. A constitutive end is an end that is constitutive of having a certain practical identity. If the agent does not adopt any means towards such ends and fails consistently to pursue them, then the agent soon stops having this practical identity. For example, if a for-profit corporation does not pursue profits, it will soon cease being a for-profit corporation. If an oil company stops pursuing its oil-related goals, at some point it will cease to be an oil company.

Constitutive ends are akin to the values mentioned in historicist principles. Recall: Mele mentions values that produce a 'Luther-style' inability not to honour them. This is the inability felt by someone with strong values, even though they can physically do otherwise. Constitutive ends induce such 'felt' inabilities. Haji and Cuypers mention components of an agent's *evaluative scheme* that bypass capacities of deliberate control; constitutive ends are parts of such a scheme. And McKenna mentions *unsheddable values* acquired while bypassing critical reflection; again, constitutive ends are unsheddable

[4] Schwenkenbecher (2023) argues that the explosion caused by mining company Rio Tinto, at Juukan Gorge in Australia, resulted from poor inter-departmental communication.

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

by a simple act of will, at least within a given timeframe. The italicized clauses are, therefore, akin to constitutive ends.

Now, one might think organizations lack values that they *cannot but* act on, that they don't have *evaluative schemes*, and that their values are never *unsheddable*. After all, one might think, a practical identity is not constitutive of being an agent: in principle, a group could reflect and change the constitutive ends of its practical identity. However, some constitutive ends of organizations *are* unchangeable, at least within a timeframe. It may be practically impossible for a company that designs machines, or sells oil, to opt for a different practical identity within a given timeframe. The oil company is committed to abide by its contracts, to follow through on its investments in infrastructure, and to the ongoing employment of specialized staff. These things take time to undo. They constitute a practical identity. They are—at least for some time period—akin to Luther-style inabilities, evaluative schemes, and unsheddable values.

Yet organizations' constitutive ends are different from their decision-making procedures and organizational structures. In Safety and Disaster, the procedures and structures (respectively) funnel the organization towards an act-*token*. With constitutive ends, the organization may not be manipulated into performing an act-token, but rather guided with respect to an act-*type*. This is because constitutive ends function as non-moral normative standards the agent evokes in assessing reasons for action. The organization is under enormous internal pressure to pursue means to those constitutive ends. The means may require the repeated performance of an act-type. While an organization could resist performing one act-token of the act-type, it is compelled to perform at least some tokens of that act-type. The question becomes: given that an organization has a constitutive end, is it reasonable to expect it to perform *zero* tokens of some act-type? If not, it has a Luther-style inability, or an evaluative scheme, or an unsheddable value, with respect to that act-type.

To illustrate, consider an 'original design' example:

> **Minerals:** At t_1, a company is created with the constitutive end of making a profit by producing phones. At t_2, the company produces phones. However, the company cannot both make a profit and constantly avoid using conflict minerals. It could avoid this for one or two niche types of phone, but not for all phones it produces. It is, practically speaking, infeasible for the corporation to change its constitutive end between t_1 and t_2, because of its contracts, investments, and employments. At t_2, the company uses conflict minerals and becomes complicit in the funding of violence and human rights abuses.[5]

The phone company can refrain from using conflict minerals in any instance. However, it cannot make a profit and refrain from using conflict minerals in *every* instance. It cannot make a profit and refrain entirely from the

[5] The role of profit in motivating the use of conflict minerals has been documented by, for example, Global Witness (Alley 2022).

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

act-type. The company is under enormous normative pressure, given its constitutive ends (and the role of profit-making within those ends), to rely on conflict minerals. Yet the company seems responsible for its complicity in human rights violations. The company could technically decide to refrain from using conflict minerals in each specific instance, or to reorientate its practical identity away from the profit motive. And surely it must know that using conflict minerals is morally wrong. Unlike with Beth, the manipulation does not seem so complete that it fully exculpates the company—as we will explain in Section IV.3.

Like Disaster, Minerals challenges historicist principles that deny the responsibility of newly formed agents, such as McKenna's Principle. The company's unsheddable end (of making profit by selling phones) did not arise from a history in which *the company* had the opportunity to critically assess, endorse, and sustain that end. The organization was set up with the end in place.

Finally, consider a radical reversal case involving constitutive ends:

> **Labor**: At t_1, a clothing company is created to make clothes ethically, with garment workers receiving fair wages and benefits. At t_2, the company starts producing clothes at a loss and continues for a few years. The clothes are popular, so the company looks good for investors. At t_3, the owners start publicly trading their shares. This is a decision made by each owner as a private individual, though each hopes the company's ethical reputation will entice ethical shareholders. The company itself is not consulted by the owners when the owners decide to sell (nor is it required to be). By t_4, the company has acquired legal responsibilities to shareholders to maximize profits. The company has good reason to believe that if it fails to maximize profit, then the shareholders will abandon the company. This would cause significant harm to the local community, as the company provides many jobs. The company develops a Luther-style inability to resist maximizing profits for shareholders. At t_5, the company has no reasonable option but to cut production costs, which foreseeably leads to sweatshop labor being used in the supply chain.[6]

Like Minerals, Labor involves constitutive ends that mandate an act-type (sweatshop labor), even though each token of that type could be avoided. The company might retain one or two ethical clothing lines—but it cannot avoid sweatshop labor entirely, consistent with maximizing shareholder profit.

In Labor, the share-selling owners stand to the company as the dean stands to Beth (although these agents have different intentions). To see this, note that a company is an agent in its own right, not merely a collection of individuals: the company is not to be identified with the owners or employees, or with any collection of individuals that includes the owners or employees. The owners and employees can act on the company 'from the outside'—without consulting the company's decision-making procedures or role structures—by

[6] The Body Shop is a real-life company whose values changed once it acquired legal obligations to shareholders (Bakan 2004).

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

Table 1: Verdicts on Examples

	Mele	Haji and Cuypers	McKenna	Judgement
Beth	Not responsible	Not responsible	Not responsible	Not responsible
Suzie	Responsible	Responsible	Not responsible	Responsible
Safety	Not responsible	Not responsible	Not responsible	Responsible
Disaster	Responsible	Responsible	Not responsible	Responsible
Minerals	Responsible	Responsible	Not responsible	Responsible
Labor	Not responsible	Not responsible	Not responsible	Responsible

selling shares or quitting the company. The only actions of owners or employees that are properly attributable to the company are those authorized by the company's decision-making procedures and role structures (Hess 2018). The share-sellers' actions are not like this. Therefore, the share-sellers manipulate the company, much as the dean manipulates Beth.

Again, in Section IV.1, we consider whether any individuals are responsible in Labor. However, again: there is clear organizational moral wrongdoing; the company must have known sweatshop labor is morally wrong; it could strictly speaking have decided not to use sweatshop labor in any given case; and there is the physical (though perhaps not volitional) possibility for it to violate its obligations to shareholders. Therefore, we assume historicists may want to hold *the company* responsible. Certainly, we could expect activists to hold it to account. Should the company really be excused just because of its newly acquired constitutive end? Again, we do not think the manipulation is sufficient to justify this—a point we defend in Section IV.3.

Yet the company satisfies Mele's Principle of non-responsibility: for years, the company's system of values precluded sweatshop labor and the company was responsible for this; the manipulation of being publicly traded was something for which the company wasn't morally responsible and to which it did not consent, yet the sale transformed its values suddenly and radically, producing a 'Luther-style' inability to do other than maximize profit. Likewise for Haji and Cuypers' Principle: the sweatshop labor resulted from the company's evolved evaluative scheme, where that evolved scheme was caused by bypassing the company's capacities of deliberative control: the company was not consulted on the sale of shares. The sale installed 'maximize profit' as a new value. According to McKenna's Principle, the company isn't responsible: it couldn't critically assess or endorse the value of maximizing profit from which it acted, yet that value was, within the relevant timeframe, unsheddable.

In all four examples, organizations challenge existing historicist principles. Table 1 states what the historicist principles say about our examples, plus the judgement we suspect many historicists will nonetheless want to endorse. Safety and Labor contest all four historicist principles. Disaster and Minerals

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

contest only those historicist principles that deny responsibility in original design cases (like Suzie). Insofar as the judgement in Disaster and Minerals (that the group is responsible) is on firmer epistemological ground than the judgement that Suzie is responsible, we conclude that Disaster and Minerals are stronger challenges to historicism than is Suzie (over whom much ink has been spilled in the historicist literature).

This points to an important choice point in the philosophy of responsibility.

IV. A TRILEMMA

IV.1 Rejecting group responsibility

In our view, group responsibility is a non-reducible and non-redundant level of responsibility. Whether *any* moral agent, individual or collective, is morally responsible for a morally wrongful action depends on whether the agent satisfies the conditions for moral responsibility (that is, the control condition, epistemic condition, and agency condition, however formulated). For group agents, this may but need not coincide with members being responsible for the same action, just as for vicarious responsibility between individuals.

However, historicists may reject that in our cases the group agent is responsible, or they may reject group responsibility altogether. They could point to other responsible agents, whose responsibility makes the organization's responsibility redundant. In Disaster and Minerals, they might hold the designers responsible. In the radical reversal cases (Safety and Labor), they might hold responsible those who made the change of values possible (the administrator in Safety and owners in Labor) or those whose presence directly precipitated the change of values (the committee members in Safety and shareholders in Labor).

We have framed the cases to minimize individuals' responsibility. Consider Labor. One could blame the owners who sell their shares, but they might be financially reliant on good returns. One could blame the new shareholders, but the company's legal obligations to shareholders are hardly their fault. And perhaps the new shareholders would be happy with meagre profits, though the company has no way of knowing this. One could assert that the new shareholders have moral obligations to be 'activist shareholders,' using their position to produce corporate good. If they failed in these obligations, then perhaps they (not the company) are responsible. However, if 'corporate good' means 'no sweatshop labor,' then the corporation would go out of business via corporate good. The shareholders would lose their investments and harm the local community's interests. It seems overly demanding to blame the shareholders for not inducing this.

One might blame whichever manager decided to use sweatshop labor. However, as theorized by List and Pettit (2011) and French (1984), the

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

sweatshop labor decision might result from procedures or structures that prevent focused deliberation on the question of sweatshop labor: perhaps different managers each made small cost-cutting decisions within their departments, which cumulated in cuts that gave implementers no option but to (unknowingly) purchase from a factory using sweatshop labor. Perhaps procedures and structures precluded implementers from questioning managers' respective cuts. Perhaps the organization's designers couldn't have foreseen this result of the procedures and structures they put in place. In short: individuals cannot always know a calamity might occur, or it might be too much to ask them to prevent it.

Likewise for the other three cases. In Safety, the decision was unintended by members, and plausibly the administrator couldn't foresee the consequences of the decision-making procedure. In Disaster, no department was responsible for causing harm, and plausibly the designers couldn't foresee how quickly department-level decisions would have to be made. In Minerals, the decision to use conflict minerals might have been unintended by any member, with designers unable to foresee that conflict minerals were needed for production. Thus, for each case, the details can be filled in such that the designers, influencers, and members are not (sufficiently) responsible.

The historicist might respond: then the calamity is a tragedy. If the organization fails to meet the historicist conditions, and if there are not (enough) other agents with responsibility, then we face a 'responsibility gap' (Collins 2019). Responsibility gaps arise when principles do not produce as much responsibility as intuitive judgements suggest. But perhaps sometimes these intuitive judgements are simply false.

However, organizations often *are held* responsible in cases like ours. This is embedded within widespread and engrained social and political practices of responsibility-holding. Our philosophical theorizing about responsibility should reject deeply embedded practices only as a last resort. This last resort need not be taken. In Section IV.3, we defend the responsibility in our examples, arguing for extra historicist principles, which outline conditions our organizations fail to meet.

Furthermore, a slippery slope arises if one views our cases as tragedies. Cyr (2020) has argued there is no relevant difference between agents who have been manipulated in certain ways (like Labor or Beth) and agents who are *constitutively unlucky*. A constitutively unlucky agent is unlucky regarding their acquired dispositions and capacities (Nagel 1976). If our cases are tragedies, so too, it seems, for all cases of constitutive luck. That verdict abandons many ordinary judgements about responsibility. We discuss this in the next section. In Section IV.3, we aim to resist this slippery slope, attributing responsibility to the organizations in our examples, to many people who are constitutively unlucky, and to Suzie—but not Beth.

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

IV.2 Endorsing structuralism and constitutive moral luck

Rather than embracing historicism and jettisoning group responsibility, we might embrace group responsibility and jettison historicism. Here, our examples are used as grist to structuralists' mill.

Indeed, those who endorse constitutive moral luck might already favour structuralism. Again, constitutive luck is luck in one's acquired dispositions and capacities. Constitutive *moral* luck arises when an agent's dispositions or capacities are not voluntarily acquired or possessed, but positively or negatively affect an agent's moral praiseworthiness or blameworthiness (Hartman 2019: 3181). If we suppose agents who are entirely constitutively lucky can be morally responsible, doesn't that refute historicism? In which case, why care that group responsibility challenges historicism? We can reject historicism, while embracing structuralism, constitutive moral luck, and group responsibility.

Indeed, consider classic arguments for constitutive moral luck. Hartman imagines that '[t]wo citizens would freely help a beggar if they had a good upbringing, but they were habituated differently. The citizen with good habituation stops to help the beggar, and the citizen with bad habituation ignores the beggar' (2018: 169). Many judge the well-habituated individual to be more praiseworthy than the badly habituated individual. This is constitutive moral luck. Perhaps, Hartman says, it's even *non-sensical* to deny constitutive moral luck: asking what someone would do without constitutive luck is like asking what they would do if they were a completely different agent. Arguably, the answer to that question cannot inform us about the responsibility of the agent as we actually find them (Hartman 2019: 3188). We cannot assess agents as they are, without assessing traits regarding which they faced constitutive luck.

Yet Hartman can't capture our examples. Hartman distinguishes two kinds of constitutive luck. First, 'responsibility-enabling constitutive luck', which 'furnishes its agent with the broad range of reason-giving cognitive abilities and reason-responsive volitional abilities required to have … "reflective self-control"' (2018: 178). Second, 'responsibility-undermining constitutive luck', which is 'certain kinds of constitutive mental properties outside of an agent's control (that result from severe emotional trauma, bad formative circumstances, systematic conditioning, and mental illness)' (2018: 177). However, the organizations in our examples have been subject to 'responsibility-undermining constitutive luck'. Hartman's distinction cannot produce their responsibility.

Of course, when explaining responsibility-undermining constitutive luck, Hartman lists processes particular to individuals. But the organizations have undergone relevantly similar processes. The company in Minerals suffered from *bad formative circumstances*: it was formed to have the profit drive, where conflict minerals were necessary to satiate this drive. The company in Labor suffered a corporate analogue of *severe emotional trauma*: the company was

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

exposed to situations creating the ingrained belief that immoral behaviours were necessary for its survival. How do trauma, formative circumstances, conditioning, and so on rule out individuals' responsibility, while ruling in the responsibility of our groups? Answering that question means finding the correct version of historicism, not abandoning historicism in favour of structuralism.

In response, one could embrace *all* constitutive moral luck. Perhaps all agents (individual and collective) can be responsible for actions flowing from how they are constituted. This is traditional structuralism. The problem is Beth. One could follow Cyr (2020), who argues that agents like Beth *are* morally responsible—to a very minor degree. For Cyr, responsibility hinges on *how many* opportunities the agent had, since the manipulation, to indirectly change their character. An agent bears a modicum of responsibility for actions performed immediately post-manipulation, when they have lacked opportunities to mitigate their constitutive luck. As more time passes, the manipulated agent acquires more responsibility, having had more opportunities to alter unlucky aspects of their constitution.

Like Hartman's view, Cyr's struggles with Section III's examples. Suppose the organization's decisions and outcomes occurred on the day that the 'original design' or 'radical reversal' occurred. Under Cyr's proposal, each organization would possess only a tiny modicum of responsibility for the workers' deaths, the dam collapse, the complicity in conflict minerals, and the sweatshop labor. Under Cyr's view, all our organizations are *as responsible* as Beth, assuming a similar opportunity for self-correction in all cases. We suspect many historicists will want to hold our organizations more responsible than Beth, even if it's granted to Cyr that Beth is a tiny bit responsible, and even if responsibility increases with opportunities-for-revision after manipulation.

IV.3 Revising historicism

The arguments of Sections IV.1 and IV.2 incline us towards revising historicism, facilitating responsibility of our organizations, Ann (Beth's industrious colleague), and Suzie—but not Beth.

Our version of historicism is 'negative', not 'positive': we suggest a responsible agent must *lack* a certain history. A responsible agent might have no history. But *if* it has a history, that history must *not* be of a certain kind. We make this choice because positive historicism has trouble attributing responsibility in Disaster and Minerals. We believe these cases are more troubling than Suzie (i.e. the argument for responsibility is stronger), largely because they are more realistic. In both cases, the agent starts performing responsible actions when it is created by design, before any opportunity for self-revision. Under positive historicism, all 'newly-formed' agents lack responsibility. This unpalatable result inclines us toward negative historicism.

We aim to formulate negative historicism to rule out Beth's responsibility, while ruling in group responsibility in our four cases. We suggest that two

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

conditions separate Beth from our four cases. These conditions are not the final word. We sketch them as a promising place for historicists to begin.

First, organizations can reasonably be expected to *guard against manipulation*, to a much greater extent than individuals like Beth. Manipulation is part of life for organizations: they are made of other agents, in the sense of material constitution, just as a statue is made of clay (Hess 2018; Hindriks 2012). This makes organizations liable to manipulation by the constituting agents—even though organizations do nonetheless bear agency of their own (Hess 2010; Hindriks 2008). Indeed, organizations are *inevitably* designed by other agents, who determine the decision-making procedures, organizational structures, and constitutive ends. Those features are then open to further manipulation by other agents. By contrast, humans are not 'designed' in this way (human parents simply lack that power over their children). It's not reasonable to expect a philosophy professor to guard against brainwashing dictated by an overzealous dean. Things might be different if the philosopher knew about the dean's brainwashing propensities—but our point is precisely that most deans do not have such propensities. Organizations can be expected to guard against manipulation: they are made of other agents and therefore are more prone to it.

This suggests the following negative objectivist historicist principle:

> **Guarding Principle.** For an agent to be morally responsible for an action to any degree, the agent must not have a history in which: the agent did all that they could reasonably have been expected to do to avoid manipulation, yet nonetheless was manipulated, which was the direct cause of the action in question.

We suggest that the Guarding Principle applies to individuals and collectives alike. It has been overlooked because individuals tend easily to satisfy it: individuals are not usually expected to do much to guard against manipulation. Organizations can reasonably be expected to do much more.

This raises a question: Does this difference between organizations and individuals imply that we should have entirely different responsibility principles for organizations than for individuals? If so, our project of finding a unified set of principles would be a fool's errand. However, notice that there is wide variation in *individuals'* propensities to manipulation. Naïve users of social media might be more prone to manipulation than those who oversee social media, for example. Such variation does not lead philosophers to apply different responsibility principles to different individuals. Likewise, we suggest, for the difference between individuals and organizations: although the latter are more prone to manipulation, this does not make them a fundamentally different kind of moral agent.

What do we mean by 'manipulation' in the Guarding Principle and throughout our examples? Following Joseph Raz (1988), we distinguish between coercion and manipulation. Both coercion and manipulation subject

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

the will of one agent to that of another. But where coercion diminishes an agent's options via external threats or pressure, manipulation distorts the way in which the agent reaches decisions, forms preferences, or adopts values and goals (Raz 1988: 378). What distinguishes manipulation from other forms of influence (e.g., advice or arguments) is that manipulation subverts and fails to respect the agency of the target: it distorts, bypasses, or supercedes the target's capacities for practical reasoning and decision-making broadly conceived (see also Fischer 2004). The more the resulting mental or physical behaviour differs from the 'baseline' of the target and the stronger the means taken to manipulate the target, the higher the degree of manipulation.

Whether it is reasonable to expect an agent to have guarded against manipulation depends on (at least) four interrelated factors: (1) the context in which the manipulation takes place; (2) the target's evidence concerning the likelihood of manipulation; (3) the means of manipulation; and (4) the target's evidence concerning the vulnerability of aspects of their agency to manipulation within this context. For example, Beth (presumably) has no evidence that indicates the dean's interference is likely nor that her value system is especially vulnerable to manipulation within the university context. But even if Beth had some evidence that she could expect *some* manipulative behaviour from her colleagues, the means taken by the overzealous dean (brainwashing) are so extreme within the workplace context that it is nonetheless unreasonable to expect her to guard herself against such forms of manipulation.

The Guarding Principle is an improvement on Haji and Cuypers' Principle. Recall: they would deny responsibility in Safety and Labor, since the organization's evolved evaluative scheme was caused by its earlier evaluative scheme while bypassing the organization's capacities of deliberate control. However, these organizations had the opportunity to guard against being bypassed. In Safety, the organization could have explicitly enshrined worker safety in its constitution, such that no committee decision could overrule that commitment. This is reasonable to expect from the organization, because (1) the organization operates in a business context where profit motives routinely place pressure on various ethical commitments, (2) the organization (plausibly) has evidence that there are various parties with strong interests to interfere with its decision-making process, meaning the likelihood for manipulation is relatively high; (3) possible attempts to change the decision-making procedures that impact its ethical commitments are not measures beyond the pale of expectation in this context; and (4) the organization (plausibly) has evidence that costly ethical commitments are likely to be the target of manipulation in this context.

Note that the specific form of manipulation (e.g. in Safety, involuntary administration and changes to decision-making procedures by the administrator) needn't be the most foreseeable form of manipulation, for it to be reasonable to expect an agent to have guarded itself against manipulation. What matters

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

is that the agent finds themselves in a context where manipulation is sufficiently likely such that it becomes reasonable to guard aspects of their agency (e.g., its ethical commitment to worker safety) against various forms of manipulation. Likewise, in Labor, the organization failed to protect their ethical production process by enshrining this into their constitution, yet it seems likely that the evidence of the organization indicates that this commitment is prone to manipulation in this commercial context. These organizations do not satisfy Haji and Cuypers' Principle for responsibility, but they also do not satisfy the Guarding Principle for being non-responsible.

By contrast, consider a variant of Labor where the company satisfies the Guarding Principle, for example by enshrining ethical production in its constitution. In such a case, the shareholders have a moral obligation to accept whatever costs arose from the company's refusal to use sweatshop labor. The organization is obliged to maximize profits only within the bounds of its constitution. Thus, we could blame shareholders if they demanded profit-maximization via sweatshop labor. Absent an explicit demand from shareholders, it would be strange for the company to believe itself at risk of losing shareholders if it did not engage sweatshop labor. So, if the company meets the Guarding Principle, yet uses sweatshop labor, either the shareholders are at fault (for demanding this practice when they knew what the constitution said upon buying shares) or the implementers of the decision would be at fault (because surely one should question the cumulative effects of managers' decisions, if those effects go against constitutionally-enshrined policies). However, the company *itself* would have done everything it reasonably speaking could be expected to do (pending the second condition we introduce below).

But the Guarding Principle is not exhaustive. Consider that Mele's Principle includes something like the Guarding Principle. The crux of Mele's Principle is that a non-responsible agent has a history of being subject to a 'very recent manipulation to which he did not consent and for which he is not morally responsible, [by means of which] his system of values was suddenly and radically transformed' (Mele 2019: 66–7). Mele could assert that a 'manipulation … for which [the agent] is not morally responsible arises only when the agent satisfies the Guarding Principle. Thus, if the Guarding Principle were all that mattered, then Mele's Principle would be sufficient.

Mele's Principle falls short in a different way: it requires that our organizations had their 'system of values … suddenly and radically transformed … [which] ensures … [they have] at least a Luther-style inability' to do otherwise than compromise worker safety or rely on sweatshop labor. Mele's Principle does not accommodate the following fact: organizations make decisions about *how* to do that which their procedures and values press them into doing. In our cases, the manipulation imposes formal or substantive features on the organization. But once imposed, these features are the organization's own.

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

The point is this. In many manipulation cases (such as Safety, Disaster, Minerals, and Labor), the precise way in which the agent pursues or responds to the manipulated features is *not set in stone*. The agent goes through a process to decide what steps to take given the manipulated features. This process facilitates the agent's responsibility. This will not happen if the agent has every aspect of their decision-making procedures, role structures, and constitutive ends entirely altered during manipulation (including features that concern how to pursue ends). In those cases, the precise way the agent responds to the manipulated features *is* determined by the manipulation and the agent is not responsible.

The previous paragraph concerns both individual and group agents. We believe organizations have a particular tendency, though: they have explicit and formalized procedures for deciding how to act, given their manipulated features (whether procedural features, structural features, or constitutive ends). Thus, organizations illustrate the deliberative gap that can arise between (1) a manipulated feature and (2) the agent's implementation of, or pursuit of, that feature. Individuals tend not to be so regimented and explicit in their decisions about how to implement and pursue their manipulated features. Groups, therefore, illustrate the need for our below Manipulation Principle—even though that principle applies to individuals too.

The suggestion is not that the organization is responsible for the imposition of the formal or substantive features. In Safety, the organization is not responsible for the imposition of a new premise-based decision-making procedure. In Disaster, the organization is not responsible for its poorly functioning interdepartmental communication structure. In Minerals, the organization is not responsible for its initial constitutive end. And in Labor, the organization is not responsible for the imposition of a new constitutive end of producing shareholder profit. But *once* these formal or substantive features exist, the organization *is* responsible for downstream choices—such as compromising worker safety (in Safety) or using sweatshop labor (in Labor).

The idea, then, is that there is an *intermediary deliberative step* between the manipulation and the action for which the organization is responsible. The group's action is not manipulated directly. Instead, what is manipulated is the group's formal or substantive features. These features are then used to produce an action that the organization cannot volitionally resist performing—where the *manner* of performance is left open. Although very little time may pass during this process, and although the organization lacks the opportunity to reverse the manipulation, the indirect nature of the manipulation of action renders the organization highly responsible for the action (contra Cyr).

The result is our second proposed principle:

Manipulation Principle. For an agent to be morally responsible for an action to any degree, the agent must not have a history in which: the agent underwent manipulation

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

> that bypassed its capacities of deliberative control, where that manipulation imposed formal or substantive features on the agent, and these features *determined the manner in which* the agent performed the action in question.

We mean 'manner in which the agent performed the action' to include the means taken to the action, as well as the realization of the action itself. We take it that Beth's manipulation is so complete that she fails to satisfy the Manipulation Principle. Yet Safety, Disaster, Minerals, and Labor are different: once the premise-based procedure (in Safety), the communication structure (in Disaster), the constitutive end (in Minerals), and shareholder obligations (in Labor) are imposed, there is an open question about *how* the organization will act in response—for example, which precise garment factories will be used, how to build the mine, how to produce the phones, and how much the company will pay for factories' products. The organizations go through explicit and formal deliberation about what exactly to do, given these manipulations.

Like Cyr's opportunity-indexed view, our Manipulation Principle allows that responsibility comes in degrees. The larger the role manipulation plays in determining how the agent performs an action, the less responsible the agent is for that action. Likewise for our Guarding Principle: the closer the agent came to doing all that they could to avoid manipulation, the less responsible they are for actions that result from manipulation. Our view differs from Cyr's in that the amount of *discretion*—rather than *opportunities for reversal*—is the crucial determinant of an agent's responsibility. On our two conditions, Beth is not responsible, not even to a minor degree.

Of course, an agent's opportunities to reverse the deliberation may *also* affect her degree of responsibility. We have not provided a full historicist account of responsibility. We have proposed two principles as addendums to existing versions of historicism, to enable those theories to handle organizations in general, and 's examples specifically. Our main contribution has been to demonstrate the important role that group-based examples can play in generating historicist principles.

V. CONCLUSION

We have provided four examples that make trouble for existing historical principles of responsibility. The result was a trilemma: reject group responsibility, reject historicism, or revise historicism. We favour the third option. To make progress on that option, we suggested two new principles—the Guarding Principle and the Manipulation Principle. We suggested that all responsible agents must satisfy these principles. The result is not that organizations are always responsible post-manipulation. But organizations do retain responsibility post-manipulation more readily than humans, since they have stronger obligations

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

to guard against manipulation and because they more explicitly consider the means to their ends. It is possible to endorse both group responsibility and historicism, though possibly more is to be learned from cases involving group responsibility than we have discussed here.[7]

FUNDING

Niels de Haan's research for this article has been funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 740,922, ERC Advanced Grant 'The Normative and Moral Foundations of Group Agency'. Stephanie Collins' research for this article has been funded by an Australian Research Council Discovery Early Career Researcher Award.

REFERENCES

Alley, P. (2022) *Very Bad People: The Inside Story of the Fight Against the World's Network of Corruption*. London: Octopus Publishing.

Bakan, J. (2004) *The Corporation: The Pathological Pursuit of Profit and Power*. New York: Free Press.

Björnsson, G. and Hess, K. (2017) 'Corporate Crocodile Tears? On the Reactive Attitudes of Corporate Agents', *Philosophy & Phenomenological Research*, 94: 273–98.

Chant, S. R. (2021) 'Responsibility Unincorporated: Group Agents and Corporate Persons', in T. Marques and C. Valentini (eds) *Collective Action, Philosophy and Law*, 176–92. Oxon: Routledge.

Collins, S. (2019) 'Collective Responsibility Gaps', *Journal of Business Ethics*, 154: 943–54.

Cyr, T. (2020) 'Manipulation and Constitutive Luck', *Philosophical Studies*, 177: 2381–94.

Fischer, J. M. (2004) 'Responsibility and Manipulation', *The Journal of Ethics*, 8: 145–77.

Fischer, J. M. and Ravizza, M. (1998) *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.

Frankfurt, H. (1988) *The Importance of What We Care About*. Cambridge: CUP.

French, P. (1984) *Collective and Corporate*. New York: Columbia University Press.

Graham, K. (2002) *Practical Reasoning in a Social World*. Cambridge: CUP.

Gilbert, M. (2014) *Joint Commitment: How We Make the Social World*. Oxford: OUP.

Haji, I. (2006) 'On the Ultimate Responsibility of Collectives', *Midwest Studies in Philosophy*, 30: 292–308.

——— (2013) 'Historicism, Non-historicism, or a Mix?', *The Journal of Ethics*, 17: 185–204.

Haji, I. and Cuypers, S. (2007) 'Magical Agents, Global Induction, and the Internalism /Externalism Debate', *Australasian Journal of Philosophy*, 85: 343–71.

Hartman, R. (2018) 'Constitutive Moral Luck and Strawson's Argument for the Impossibility of Moral Responsibility', *Journal of the American Philosophy Association*, 4: 165–83.

——— (2019) 'Moral Luck and the Unfairness of Morality', *Philosophical Studies*, 176: 3179–97.

Hess, K. (2010) 'The Modern Corporation as Moral Agent: The Capacity for 'Thought' and a "First-Person Perspective"', *Southwest Philosophy Review*, 26: 61–9.

——— (2018) 'The Peculiar Unity of Corporate Agents', in K. Hess, V. Ignesky and T. Isaacs (eds.) *Collectivity: Ontology, Ethics and Social Justice*, 35–60. London and New York: Rowman and Littlefield.

[7] We would like to thank members of the ERC project 'The Normative and Moral Foundations of Group Agency' at the University of Vienna and two anonymous referees for helpful comments.

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023

Hindriks, F. (2008) 'The Status Account of Corporate Agents', in H.B. Schmid, K. Schulte-Ostermann and N. Psarros (eds) *Concepts of Sharedness: New Essays on Collective Intentionality*, 119–44. Frankfurt: Ontos Verlag.

——— (2012) 'But Where Is the University?', *Dialectica*, 66: 93–113.

——— (2018) 'Collective Agency: Moral and Amoral', *Dialectica*, 72: 3–23.

Isaacs, T. (2011) *Moral Responsibility in Collective Contexts*. New York: Oxford University Press.

Levy, N. and McKenna, M. (2009) 'Recent Work on Free Will and Moral Responsibility', *Philosophy Compass*, 4: 96–133.

List, C. and Pettit, P. (2011) *Group Agency: The Possibility, Design and Status of Corporate Agents*. Oxford: OUP.

McKenna, M. (2004) 'Responsibility and Globally Manipulated Agents', *Philosophical Topics*, 32: 169–92.

——— (2006) 'Collective Responsibility and an Agent Meaning Theory', *Midwest Studies in Philosophy*, 30: 16–34.

——— (2016) 'A Modest Historical Theory of Moral Responsibility', *The Journal of Ethics*, 20: 83–105.

Mele, A. R. (2006) *Free Will and Luck*. New York: Oxford University Press.

——— (2016) 'Moral Responsibility: Radical Reversals and Original Designs', *The Journal of Ethics*, 20: 69–82.

——— (2019) *Manipulated Agents: A Window to Moral Responsibility*. New York: Oxford University Press.

Nagel, T. (1976) 'Moral Luck', *Proceedings of the Aristotelian Society*, Supplementary Volumes (50): 137–51.

Raz, J. (1988) *The Morality of Freedom*. Oxford: OUP.

Rovane, C. (1998) *Bounds of Agency*. New Jersey: Princeton University Press.

Schwenkenbecher, A. (2023) 'Do Group Agents Resemble Psychopaths?', Unpublished Manuscript.

Vargas, M. (2006) 'On the Importance of History for Responsible Agency', *Philosophical Studies*, 127: 351–82.

Watson, G. (2004) *Agency and Answerability: Selected Essays*. New York: Oxford University Press.

Wolf, S. (1987) 'Sanity and the Metaphysics of Responsibility', in F. Schoeman (ed.) *Responsibility, Character, and the Emotions*, 51–69. Cambridge: CUP.

[1] *Department of Philosophy, Monash University, Australia*

[2] *Department of Philosophy, University of Vienna, Austria; African Centre for Epistemology and Philosophy of Science, University of Johannesburg, South Africa*

© The Author(s) 2023. Published by Oxford University Press on behalf of The Scots Philosophical Association and the University of St Andrews. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/pq/advance-article/doi/10.1093/pq/pqad104/7329479 by guest on 26 October 2023