Received: 28 May 2024 | Revised: 11 September 2024 | Accepted: 21 September 2024

DOI: 10.1111/phc3.70002

ARTICLE

WILEY

Responsibility Gaps

Michael Da Silva

University of Southampton, Southampton, UK

Correspondence
Michael Da Silva.
Email: M.da-silva@soton.ac.uk

Abstract

Responsibility gaps arise when there is a mismatch between the amount of responsibility that can be attributed to any person or collection of persons on leading accounts of moral responsibility and the amount that robust intuitions suggest should be allocated to someone in a case. Claimed responsibility gaps arise in numerous philosophical debates, including those concerning government, corporate, and other forms of group agency and new technologies and those concerning theoretical issues in the philosophy of responsibility. This work is an opinionated introduction to and overview of recent work on responsibility gaps. It outlines and evaluates paradigmatic responsibility gap cases and ways of understanding the phenomenon as well as the existence conditions and moral status of and possible responses to responsibility gaps. It thereby contributes to ongoing work in the philosophy of responsibility and several applied domains.

1 | INTRODUCTION

Responsibility gaps arise when there is a mismatch between the amount of responsibility that can be attributed to any person or collection of persons on leading accounts of moral responsibility and the amount that robust intuitions suggest should be allocated to someone in a case. There is, in other words, "a deficit" in the moral "accounting books" (Pettit, 2007, p. 194) arising from how "traditional ways of responsibility ascription are not compatible with our sense of justice" (Matthias, 2004, p. 177). These can create situations whereby no one is accountable for harms that arise or make it difficult to seek compensation or even apologies or explanations for the harms.[1]

Many scholars seek ways to 'fill' responsibility gaps and address these problematic deficiencies in fundamental interests in, for example, accountability and answerability.[2] Others deny that responsibility gaps exist or call for some response.[3] However, appeals to responsibility gaps appear in several philosophical domains, including prominent debates about applied issues concerning government, corporate, and other forms of group agency and new technologies and theoretical issues in the philosophy of responsibility.[4] Understanding the nature of and controversies concerning responsibility gaps can help one navigate such debates. This article accordingly provides an opinionated overview of recent scholarship on responsibility gaps.

2 | (POSSIBLE) RESPONSIBILITY GAP CASES

Cases are central to and help motivate responsibility gap scholarship. Many prominent ones involve large human-caused disasters. Pettit (2007, p. 171), for example, offers the following:

> [A] ferry operating in the English Channel, sank on March 6, 1987, drowning nearly two hundred people. The official inquiry found that the company running the ferry was extremely sloppy, with poor routines of checking and management. … But the courts did not penalize anyone in what might seem to be an appropriate measure, failing to identify individuals in the company or on the ship itself who were seriously enough at fault.

While Moen (2024) suggests this case is non-exemplary because standard gap cases involve groups with a clear organizational structure, crashes are central to several prominent works on gaps. For instance, Himmelreich (2019) appeals to the Exxon Valdez crash/oil-spill as a possible gap case and Collins (2019) appeals to the Mount Erebus plane crash in the related collective duty gap context. Each case involves individuals who are responsible for the crash and its effects but whose joint responsibility on leading theories of responsibility is insufficient to address all harms. Robust intuitions about the amount of responsibility that should be attributed somewhere remain unfulfilled – and so too understandable desires for one to face sanction, apologize, etc.

Other prominent cases stem from group decision-making procedures. Braham and Van Hees (2011, p. 11) helpfully summarize another case in Pettit that Moen (2024) considers exemplary:

> A three-member committee of an employee-owned firm has to decide whether or not to impose a pay-cut in order to finance new workplace safety measures. The decision to impose the pay-cut and implement the safety measure is to be taken if and only if an affirmative verdict is reached on three issues, each of which is determined by simple majority. [Each member votes against at least one issue and so would oppose the pay cut. A majority votes positively for each issue separately. The pay cut is implemented.]

This case exemplifies the 'discursive dilemma' (List, 2006; Pettit, 2001) or 'doctrinal paradox' (Chapman, 1998; Kornhauser & Sager, 1993) whereby the aggregation of individual judgments on a set of criteria produces a conflict between the overall outcome desired by each individual decision-maker and the final decision reached. Other examples include a committee determining whether a professor fulfills three criteria for tenure (Braham and Van Hees, 2011; Copp, 2006; Pettit, 2007) and a three-member jury assessing whether an accused person is guilty of a crime (Kornhauser & Sager, 1993; Moen, 2024). These cases share a structure: each voter can point to their consistent votes to suggest they are not responsible for the collective outcome. Responsibility gaps may arise where, for example, employees feel wronged by the pay cut and yet cannot find any person (or group of persons) who they can aptly blame, ask for compensation, etc. on traditional accounts of responsibility.

Still other cases focus on technological harms. Appeals to responsibility gaps are common in discussions of 'killer robots' (Sparrow, 2007). Scholars worry that autonomous warfare systems (AWS), like drones, may produce decisions that cannot be fully attributed to military personnel or developers, creating gaps.[5] While some seek to avoid apparent gaps involving AWS by holding military commanders responsible, other AI cases lack entities with relevant role responsibilities. Consider self-driving cars that make errors developers could not reasonably predict (Hindriks and Veluwenkamp, 2023). Yet another case exemplifies general worries with AI:

> [A] system for the automatic diagnosis of lung cancer […] learns to identify cancer cells on microscope images of specimens of needle biopsies obtained from the bodies of the persons to be diagnosed. … The system has been constructed so that false negative diagnoses are highly improbable … but there is accordingly much less precaution about false positives … [that] can cause great financial, practical and emotional problems.
>
> (Matthias, 2004)

Not all cases involving unaddressed harms plausibly present gaps. Consider a natural disaster for which all reasonable efforts to minimize damage were made. One can understand a person seeking someone to hold responsible when the storm destroys the person's home. Yet all parties in relevant debates accept that expecting someone to be held morally responsible would be unreasonable. This is likely due to the nature of responsibility gaps. As we will now see, responsibility gaps exist only where there is a break between the amount of responsibility that robust intuitions deem necessary to attribute and the mount traditional theories can attribute. Responsibility gaps do not arise absent robust (and, plausibly, justified) intuitions that *someone* should be held responsible for the harms at issue, which are missing in natural disaster cases.[6]

3 | THE PHENOMENON

Intuitions about cases (and the possibility of gaps) differ, but the cases above exemplify the (purported) phenomenon and provide a useful touchstone for analysis. They share a common feature that characterizes the basic phenomenon, namely a mismatch between the amount of responsibility one can attribute to any person or set thereof on leading accounts of responsibility and the amount of responsibility robust intuitions would otherwise deem necessary to attribute.

Collins (2019, p. 943) helpfully frames the issue in the collective responsibility context with reference to "holes" in "situations in which we are unable to attribute all the responsibility we might pre-reflectively want to attribute to collectives, such as business corporations and states." Theoretically-justified responsibility attributions do not match a "pre-theoretical intuition" or "gut reaction" about the responsibility quanta that should be ascribable (or who should have it) (946). Copp (2006, p. 216) likewise appeals to the need for collective agency where "there is moral or rational fault that must be assigned somewhere" but that one cannot fairly assign to collective members. Pettit (2007)'s aforementioned accounting deficit (see also Moen, 2024) partly stems from a lack of "fit" of individual, group, and total responsibility. Köhler et al. (2018, p. 54) similarly suggest gaps produce circumstances where traditional accounts lack "resources to say what should appropriately be said" about responsibility in a case. These gaps are sometimes said to apply only where "no one" can be properly held responsible for a state of affairs.[7] However, paradigmatic cases render *some* individuals responsible for aspects of an outcome.[8] The cases are problematic where and because there is a mismatch between desired and appropriately attributable responsibility. The mismatch is worrisome, not the lack of *one* responsible agent. The missing agent to whom 'no one' would refer is just one who could remedy the mismatch.

Responsibility gaps, then, arise when no entities meet all the conditions for moral responsibility to the degree sufficient to attribute responsibility for all harms that intuitively should be addressed. Most scholars frame the

deficits in light of a lack of persons who fulfill folk conditions on responsibility, like causal control, relevant knowledge, and opportunity to act otherwise.[9] Braham and Van Hees (2011) identify 'responsibility voids' corresponding to each condition, distinguishing causal, epistemic, and normative variants. Yet all responsibility gaps purportedly implicate normative considerations. Gaps are meant to address the lack of someone who can be justifiably 'held responsible' in the sense of being subject to certain forms of treatment and even duty-bound to fulfill certain obligations. Pettit (2007) distinguishes being held responsible in the sense of being an appropriate subject of blame or praise; final accountability in the sense of being "the one who carries the can … [or] sits at the desk where the buck stops;" and forms of regulating conduct whereby we subject persons to sanction to deter them from future acts or develop their moral character.[10] While gaps may also subject one to other forms of treatment below, gaps appear problematic when there is a lack of the first kind of moral responsibility, rather than of mere legal liability or policies for avoiding future harm.

Responsibility gaps are distinct from several related problems, though there are enough parallels to draw from discussions of other problems involving forms of responsibility "blurring" (Köhler et al., 2018). Collins (2017a, pp. 573–574), for example, contrasts collective duty gaps and collective responsibility gaps: "The latter are gaps in backwards-looking blame, control, agency, or causation; the former, gaps in forward-looking duty or obligation. … Duty gaps arise out of responsibility gaps, yet duty gaps are more morally tractable." They can, moreover, be filled with something other than individual-level blame, control, agency, or causation vis-à-vis the harm. One should distinguish these related phenomena. Discussion of "retribution gaps" (Buell, 2018; Danaher, 2016) are also related insofar as those gaps stem from the lack of an agent who can be subjected to desired retribution. However, the gaps at issue do not stem from the inability to fulfill retributive urges or problems holding those who are morally responsible to account. They instead stem from no entity being held morally responsible to seemingly appropriate degrees.

Discussing 'responsibility gaps' as 'responsibility voids' (Braham and Van Hees, 2011; Duijf, 2018) risks misunderstanding where a 'void' is only one of three conditions for a responsibility gap in a leading account. Per Himmelreich (2019, p. 734), a gap occurs iff "(1) a merely minimal agent does x, such that (2) no one is responsible for x; but (3) had x been the action of a human person, then this person would be responsible for x." 'Voids' only appear in (2). The first "Minimal Agency Condition" seeks to distinguish gaps from occurrences such as floods or landslides that do not appear morally problematic in the same way (735). The second "Responsibility Void" and third "Lack of Moral Agency" Conditions then create circumstances where one would seek to attribute responsibility for X to the degree Y but cannot do so. Responsibility gaps arise where one can say things like 'If I spilled oil like Exxon, I would be properly held responsible.' Himmelreich's claim to offer the first set of necessary and sufficient conditions is plausible. The conditions are hard to apply if 'gaps' and 'voids' are synonymous.

Finally, the 'Problem of Many Hands' whereby many persons contribute to an outcome such that it is "difficult even in principle to identify who is morally responsible" (Thompson, 1980) is sometimes discussed as a responsibility gap problem (van de Poel et al., 2012) or more broadly in terms of concerns holding anyone responsible (Bovens, 1998). When the Problem of Many Hands is framed in terms of many small contributions to a large outcome (e.g., Sondermann et al., 2018, p. 2), it risks producing responsibility gaps at issue here. However, the Problem of Many Hands and responsibility gaps may not perfectly coextend. It is, for instance, notable that paradigmatic instances of the Problem of Many Hands, like climate change, are said to produce duty gaps (Collins, 2019). Even if the Problem of Many Hands is best understood as including both responsibility gap and duty gap problems, it is worth distinguishing these components. Additionally, not all responsibility gaps arise from actions (or inaction) by 'many' persons. Note, for example, the three-person structure of classic discursive dilemma/doctrinal paradox cases.

4 | CONDITIONS FOR/TYPES OF GAPS

Scholars only recently began discussing existence conditions for responsibility gaps. The aforementioned Himmelreich (2019) is one example. Köhler et al. (2018, p. 54), in turn, introduced the mismatch-based view with necessity conditions requiring that "(1) it seems fitting to hold some person(s) to account for some φ to some

degree" and "(2.1) there is no candidate who it is fitting to hold to account for φ or (2.2) there are candidates who appear accountable for φ, but the extent to which it is, according to our everyday understanding, fitting to hold them individually to account does not match D." These conditions are useful even if Köhler et al.'s definition of responsibility gaps does not also clearly aim to provide sufficiency conditions for those gaps.[11]

Responsibility pluralists suggest there are many forms of 'responsibility.' This leads some to posit further 'gaps' and others to deny the phenomenon. Köhler et al.'s conditions focus on accountability. Pettit (2007), Matthias (2004), and other prominent scholars also discuss responsibility gaps in light of a lack of fitting subjects for all blame (or praise). Yet more recent analyses warn against what they see as an undue focus on accountability (e.g., Tigard, 2021, p. 599; Glavaničová & Pascucci, 2022). Pluralists suggest responsibility also implicates other fundamental interests, like answerability (Shoemaker, 2011, 2015). Responsibility is not just about allocating praise or blame but also about allocating duties to explain or apologize for (usually negative) outcomes. Shoemaker (*id*.) further suggests that answerability and accountability do not exhaust the category of persons who are properly labelled as 'responsible.'[12] One can 'attribute' an action (for example) to a person in ways that makes them fairly liable to moral appraisal even if that person is not fully answerable or accountable for the action. Acts that evince persons' character traits but are not themselves triggers for explanatory or punitive duties are meant to be examples.

Pluralism complicates discussions of responsibility gaps. On the one hand, it offers means of dissolving or filling gaps. Scholars note that failure to attend to different elements of responsibility can lead others to posit gaps unnecessarily. To wit, Tigard (2021) suggests some apparent accountability-based gaps can be addressed by other responsibility practices, like the reason-giving required by answerability. Responsibility pluralism accordingly leads some to deny genuine gaps can arise.[13] On the other hand, if multiple forms of responsibility exist, gaps may arise in multiple forms. Some gaps in one form of responsibility may not be capable of being filled by other forms. Those who believe in responsibility gaps posit gaps in accountability, answerability, or both (if not attributability). The absence of an answerability gap does not entail the absence of an accountability gap (or vice versa) on a pluralist view in which forms of responsibility do not coextend. Intuitive responsibility deficits stem from the lack of someone who can aptly apologize for or explain harms. Understanding gaps in terms of accountability alone then appears problematic.

Pluralism, in fact, raises broader questions about the forms of responsibility a plausible account of gaps must address. Some scholars offer capacious views. For example, de Sio and Mecacci (2021, pp. 1068–1069) posit four gaps: a "culpability gap" concerned with the lack of a responsible agent; a "moral accountability gap" in which persons are less able to reflect on their and explain others' behaviour; a "public accountability gap" concerned with persons being unable to get reasons for actions; and a technology-specific "active responsibility gap" concerning persons' understanding of the relationship between their actions and that of technology. One apparently misses a "broader picture" by focusing on a single gap, producing misunderstandings that hinder responses thereto. Still others believe responsibility has forward-looking aspects cases above do not address: "Prospective responsibility gaps" (Collins, 2019, p. 949) between obligations that must be fulfilled (intuitively by specific bodies) and those we can justifiably ascribe also occur.

The mismatches at issue nonetheless identify a distinct phenomenon. Treating them under a common rubric with other, related problems in de Sio and Meccaci's taxonomy risks confusion. Solutions to the other purported 'gaps' they identify would leave our mismatches in place. While an exclusive focus on accountability may miss important concerns, the mismatch-based account can accommodate that worry. Those interested in 'backwards-looking' responsibility gaps have long recognized that liability to punishment or reactive attitudes like blame do not exhaust the relevant kind of responsibility; responsibility also applies to, for example, apt calls for explanation (e.g., Matthias, 2004, p. 175). Himmelreich (2019)'s conditions can also be understood in light of concerns with answerability, if not attributability. Köhler et al. (2018) could be similarly rewritten. Kiener (2022), in fact, articulates mismatch concerns in answerability cases.

While 'responsibility' may, in turn, attach to forward-looking or public accountability and active responsibility gaps, they do not clearly constitute one "broader picture." Even pluralists admit they have different causes. Underlying concerns qualitatively differ and may not submit to similar solutions. Explainable AI requirements may, for instance, improve public accountability and yet leave a 'culpability gap' in place. Providing explanations for a state of affairs likewise cannot obviously fill a culpability gap where, for instance, the problem concerns a perceived need for redress.

While 'responsibility' and 'responsibility gap' may refer to many concepts, then, mismatches between the amount of responsibility that can be attributed to any person or collection of persons on leading accounts of moral responsibility and the amount of responsibility that intuitively should be allocated to someone in a case are a distinct phenomenon. Even responsibility pluralists should distinguish these responsibility gaps from nearby phenomena. Gaps so-defined could refer to the lack not only of one to blame but also of one to provide explanations, someone to provide an apt apology, or otherwise fulfill the intended functions of any theory of moral responsibility. One's preferred theory will specify those functions and thus possible gaps.

5 | THE MORAL STATUS OF GAPS

Many scholars who believe responsibility gaps exist suggest that such gaps are problematic because they leave persons unable to hold anyone accountable for harms imposed upon them or receive reasons, apologies, compensation, or redress for the harms (e.g., Köhler et al., 2018; List, 2021; Pettit, 2007; Sparrow, 2007). If responsibility gaps are defined in terms of a lack of someone able to perform these functions, then they have at least one bad-making feature. Other scholars suggest that responsibility gaps only occur where there is a *problematic* mismatch between desired responsibility attributions and those available on leading accounts (e.g., Matthias, 2004; van de Poel et al., 2012). Gaps are accordingly morally suspect by definition.

Recent work challenges the view that gaps must be problematic. Danaher (2022), for example, contends that gaps can be desirable. Leaving gaps in place can, Danaher suggests, provide the best response to moral dilemmas where even rightful actions will produce harms that intuitively call for a response. Danaher contrasts three possibilities for how to address dilemmas: 'Delegation,' in which we "get someone (or some *thing*) to make tragic choices on our behalf"; 'Illusionism,' which denies tragic choices exist; and 'Responsibilisation,' in which persons just bear the choices' costs. According to Danaher, Delegation is sometimes preferable to alternatives even where it leaves gaps in place. In particular, Danaher suggests, delegating difficult decisions to machines can be desirable even where the machine itself cannot be held responsible and it would be inappropriate to hold someone else responsible for its actions. Such delegation is, Danaher suggests, psychologically attractive, shifts relevant burdens to those better able to bear costs (particularly in machine-based cases where they cannot experience choice-based angst), and can fulfill social benefits by getting people to do what should be done. This makes Delegation desirable even absent means of attributing responsibility for attendant harms to any entity or collection thereof. It can then be better to let gaps arise than to avoid the gaps or 'fill' them by finding responsible parties. Danaher believes Illusionism and Responsibilisation are also appropriate in some cases and his arguments focus on technological cases. However, he raises the possibility that letting some gaps arise – and, perhaps, remain unfilled – can be desirable.

Additional possibilities are available but rarely discussed in the literature. Gaps may, for instance, present neutral cases. 'Responsibility gaps' can be a descriptive term denoting a mismatch rather than an explicitly normative phenomenon. Gaps on this view provide an appearance of an issue in that theory and intuitions do not cohere. However, whether and when this is problematic will be independent of the existence of a gap. Relatedly, gaps could be considered as having bad-making features but not all-things-considered bad. The scope/force of the bad-making feature and how to respond to gaps will then depend on their (real or potential) consequences. A responsibility gap may, for instance, itself be minimally problematic, providing reason to let it arise. But if the

continued existence of the gap is likely to leave many persons unable to receive compensation for severe harms, that will be bad and demand some response.

6 | RESPONSES

One's response to responsibility gaps will likely depend on how one characterizes them and their status. A possible response is to alter one's theory of moral responsibility. If a theory produces too many gaps, it risks becoming generally non-intuitive and/or having bad consequences. Theoretical refinement or change is a natural response. Gaps can, indeed, plausibly trigger an exercise in reflective equilibrium, seeking to balance theoretical and case-based considerations. One could, for instance, alter the control condition on responsibility to avoid producing gaps.

Most common responses deny or minimize the phenomenon or seek to fill gaps. The first set of responses includes denying their general existence or presence in a domain (e.g., Grübler, 2011; Himmelreich, 2019; Köhler et al., 2018; Tollon, 2022) and suggesting they are rare, uninteresting, or not as problematic as claimed (e.g., Braham & Van Hees, 2011; Duijf, 2018). Several scholars seek to "dissolve" (Danaher, 2022) gaps by demonstrating that all or most cases permit one to identify particular persons that have sufficient control to warrant full responsibility when properly described (Himmelreich, 2019; Moen, 2024).[14] Königs (2022), for example, suggests gaps are a "philosophical mirage": Even cases involving AI do not involve genuinely autonomous systems creating harm free from the responsible actions of their developers and users; rather, harms result from developer or user recklessness or malice. Others suggest particular personnel must bear responsibility as part of their role in a group, as when military commanders take responsibility for acts under their command.[15] A less common response accepts that responsibility gaps exist and attempts to avoid conditions producing them. Sparrow (2007), for example, called for a prohibition on the use of AWS partly to avoid responsibility gaps.

The second set of responses views gaps as genuine and calling for a response. Attempts to fill (or "plug" (Danaher, 2022)) gaps take many forms. Some posit entities who can take responsibility. These approaches too engage in reflective equilibrium between theory and practice but do not change the conditions of responsibility. They instead identify new potential subjects of responsibility. Pettit (2007, p. 194) claims corporate responsibility is "the only possible way to guard against" gaps. While group agency is controversial, it offers a means of addressing gaps in individual responsibility. Likewise, AI personhood remains controversial but offers another locus of responsibility for filling gaps. Yet challenges remain. For one, gaps may still occur where traditional accounts do not permit individual and group agents (or AI) to jointly bear full responsibility for the quantum of responsibility intuitively desired. For another, non-human agents may not be able to perform acts necessary to fill gaps. If AI cannot hold and expend funds, it cannot pay compensation. And if a corporation cannot experience pain, it cannot fill responsibility gaps that call for someone to face retribution and physical punishment. Danaher (2022)'s still-rarer response suggests one should leave some gaps in place – or even desire them.

Others suggest different moral or legal mechanisms can play roles desired by 'gap-filling' without positing new agents. Collins (2019), for one, suggests some persons should pick up the responsibility 'slack' in analogous cases with collective duty gaps. Such "solutionism" (de Sio & Mecacci, 2021, pp. 1073–1074) also includes legal (e.g., compensation schemes, AI personhood, and no-fault regimes) and technical methods for addressing apparent gaps. For yet another example, Glavaničová and Pascucci (2022) note that ascribing responsibility can be helpful for "attributing conduct, blaming wrongdoers, imposing a duty to account or answer for what happened, punishing the guilty, preventing future harm, compensating victims," etc. and suggest that imposing vicarious liability for those with some, but not total, control over an outcome will best achieve these valuable ends. These may offer possible solutions to apparent problems with gaps.

Questions remain regarding whether such solutions genuinely 'fill' gaps. Gap-filling, recall, is meant to fix a mismatch in responsibility attributions. Someone is supposed to be held responsible for the harm, rather than someone merely being held legally liable or new rules being established to regulate future conduct (Pettit, 2007).

Regulation may address some of the functional problems caused by responsibility gaps. Pettit and others accept that legal liability could address some of those problems even if they keep gaps in place. Yet even laws proposed for that purpose should remedy the specific responsibility-related issues gaps produce. Whether and when this is possible or worthwhile are live issues. Vicarious liability imposes costs on persons who are, ex-hypothesi, not morally responsible on traditional accounts of responsibility. The costs may not be worthwhile. And legal liability may leave gaps in genuine responsibility in place. Treating people as if they were responsible may not satisfy the desire for apt attributions.

Finally, Pettit's collaborator (List & Pettit, 2011) List (2021) stakes a middle ground: One can fill gaps in AI responsibility as one does in corporate cases but if the AI itself cannot meet the conditions for responsibility, then one can require someone else to accept strict liability for AI-induced harms as a second-best solution. The latter condition also exemplifies attempts to use role responsibilities to fill, rather than dissolve, responsibility gaps. This position help to address some challenges facing Pettit. However, it could also raise questions about *how* gaps are 'filled.'

Seeking to fill (at least many) responsibility gaps is plausible if gaps exist. However, gap-filling accounts face at least two challenges hinted at above. One concerns when gaps should be filled. Several options for when to fill gaps are available. Leaving them all in place is theoretically possible, but even those who believe gaps can be desirable (Danaher, 2022) believe some should be filled. Alternatively, one could seek to fill all gaps, but deviating from the responsibility attributions justified by the best traditional theories thereof is bound to have costs that may not always be worthwhile. I suspect most will instead seek to identify particular conditions under which gaps should be filled. This too admits multiple options. One could seek to determine rules for when to fill gaps. One could, for instance, take a consequentialist approach and suggest gaps should be filled when the value of addressing the harms is greater than the costs of deviating from traditional attributions. Alternatively, one could take a contextual or pragmatic approach whereby decisions on when to fill gaps should be made on a sector-by-sector or case-by-case basis. This approach requires guidelines for how to make contextual or pragmatic decisions. Indeed, additional detail is required to specify both consequentialist and contextual responses.

The second challenge concerns what it would mean to 'fill' gaps – and whether this is possible. One can 'fill' gaps by treating someone responsible in a theory of responsibility or one can 'fill' them practically. If gap-filling is meant to be theoretical, someone must be able to aptly take on the relevant form of responsibility. However, whether one consistently can do so without dissolving gaps is unclear. Kiener's critique of appeals to role responsibilities to fill accountability gaps is representative. Per Kiener (2022), either CEOs/commanders in such cases are not blameworthy and so cannot actually accept responsibility that would trigger accountability-related harms or CEOs/commanders *are blameworthy* and so appropriately held responsible, eliminating, rather than filling, purported gaps. Similar problems may apply to other means of potentially filling gaps. It is, for instance, unclear whether a new CEO can genuinely apologize for conduct in which they played no part. If the CEO is an apt candidate for apologizing, this is due to their blameworthiness. Simply changing one's theory of responsibility to avoid this outcome will raise problems with other simple alterations outlined above. Recognizing new agents– like Pettit's corporations – could, in turn, identify parties that one could hold responsible. But it presents its own challenges above and such agents cannot always clearly 'fill' the gaps as intended. Kiener instead draws on Enoch (2012) and proposes identifying persons who can 'take responsibility' after the fact and accept the consequences that a fully responsible party would have faced. Such an approach would, plausibly, at least identify a person who could fulfill the problems that gaps produce. Yet it raises challenges concerning when and how persons can 'take' responsibility for the relevant kind of harms. Even Kiener and Enoch offer different accounts of who can 'take' responsibility when. And Enoch notably believes that one cannot choose to accept blame. This approach could leave blame gaps unfilled.

If gap-filling is merely meant to be practical, the relevant practices may leave moral responsibility gaps in place and raise their own challenges. Practical mechanisms for 'filling' gaps by deeming particular persons liable

to pay compensation, apologize, etc. can address apparent problems posed by gaps on some accounts above. However, they also risk leaving the responsibility gaps in place. If the person deemed legally liable for a harm is not genuinely responsible for it, the responsibility gap remains. If gaps as such are problematic, practical solutions cannot resolve all underlying problems. And it is not clear that anyone is genuinely held responsible on such approaches. Rather, people are rendered liable to assign final accountability or regulate conduct absent genuine moral responsibility. Whether even these practical functions can be fulfilled is still another matter. The aforementioned apology by the newly-mentioned CEO may, for instance, leave even the practical need for an apology unaddressed.

7 | CONCLUSION

Sometimes our best theories of moral responsibility do not account for robust intuitions about the need for someone to bear responsibility for an act or state of affairs. Traditional theories cannot assign the full quantum of responsibility intuitively owed to those harmed to any entity or even collection of entities. This mismatch between theory and intuition presents a responsibility gap.

Appeals to responsibility gaps appear in numerous domains. While particular cases and even the general existence of gaps remain controversial, the basic idea is influential. Many questions about responsibility gaps remain even if they exist. The preceding accordingly offered an overview of existing work on the conditions for and types and moral status of responsibility gaps and on how to respond to them (if they exist). It also identified outstanding theoretical and practical challenges, including those related to what it would mean to fill responsibility gaps.

Further areas of inquiry remain. One interesting question concerns whether the purported gaps above even refer to a common phenomenon. One may, for instance, query whether apparent gaps in corporate and AI settings refer to the same phenomenon or whether gaps in accountability and answerability should be considered under a common framework. If there is a common phenomenon, one can further question whether they submit to a common solution. Others may suggest the preceding misses phenomena that should fall under the responsibility gap umbrella. There may, for instance, be closer links between future-oriented accounts of responsibility and the primarily-backwards-looking accounts above than I recognized. This idea merits scrutiny.

Examining other possible areas of application could also prove fruitful. Work on responsibility gaps often intersects with work on state responsibility and the relationship between states and their members. A related area of research examines whether international bodies can fill apparent 'gaps' in state responsibilities. Whether the kinds of structural injustices international institutions aim to address are best characterized in terms of gaps is another possible area for future inquiry.

I cannot provide solutions to these questions in this overview. However, they combine with the theoretical and practical considerations (and challenges) above to highlight the potential significance of responsibility gaps. Even if all gaps can be dissolved, knowing why is important.

ACKNOWLEDGMENTS

The author thanks Hannah Da Silva and two anonymous reviewers for feedback on earlier drafts and Theron Pummer for shepherding the piece through review. Thanks are also due to those who read related texts or listened to related talks. Éliot Litalien, Brian McElwee, and an audience at the Zicklin Center for Governance and Business Ethics are notable in this respect.

CONFLICT OF INTEREST STATEMENT

The author declares no conflicts of interest.

ENDNOTES

[1] E.g., Pettit (2007); Sparrow (2007); Köhler et al. (2018); List (2021).

[2] Shoemaker (2011, 2015) distinguishes three elements or forms of responsibility. Answerability identifies parties who owe reasons or a justification for a decision, outcome, etc. Accountability identifies those who are appropriately subject to blame or censure (and their positive corollaries). Attributability identifies persons who are properly labelled as 'responsible.' On Shoemaker's account, this need not coextend with being answerable or accountable for an action or state of affairs since the responsibility label can make you liable to other forms of moral appraisal even where it does not trigger accountability or answerability. Whether these categories are distinct remains contested. Smith (2012), for example, argues that Shoemaker's categories can all fall under an answerability framework. It suffices here to note that scholars posit gaps in (at least) accountability and answerability. See below for examples of scholars linking responsibility gaps to concerns with accountability and answerability respectively.

[3] Again, see below for specific examples.

[4] On groups, including corporations, see, e.g., Pettit (2007); List and Pettit (2011); List (2021); Braham and van Hees (2011); Duijf (2018). Smith (2009) discusses similar phenomena. On states, see, e.g., Lawford-Smith and Collins (2017). The volume including Köhler et al. (2018) also largely focuses on states. On automated warfare systems, see, e.g., Sparrow (2007); Danaher (2016, 2022); Himmelreich (2019); Swoboda (2017); Zając (2020); Oimann (2023). On 'self-driving' cars, see, e.g., Danaher (2016); Nyholm (2017); de Jong (2020). On artificial intelligence generally, see, e.g., Matthias (2004); Champagne and Tonkens (2015); Köhler et al. (2018); Tigard (2021); de Sio and Mecacci (2021); Glavaničová and Pascucci (2022); Tollon (2022). Some authors address multiple phenomena and parallels between them.

[5] Note the example in Himmelreich (2019, p. 273).

[6] On natural disasters, see also List (2021, p. 1127). Seeking a responsible party in natural disaster cases of this form can be psychologically understandable, but remains inappropriate. Himmelreich (2019)'s necessary and sufficient conditions for responsibility gaps explain the absence of a gap here in terms of the lack of even a minimal agent. He believes gaps only occur where a minimal agent does something for which a human would be held responsible.

[7] Braham and Van Hees (2011) discuss cases where no one "can individually be held morally responsible for an outcome." Danaher (2016) frames the issue in terms of a lack of a "culpable wrongdoer." Himmelreich (2019, p. 735) suggests gaps occur where "no one can be responsible." Copp (2006), Pettit (2007), and Duijf (2018, p. 457) use "no one" language in similar contexts. Mukerji and Luetge (2014, p. 176) view this as deflationary: if "nobody is responsible" in key cases, the concept of moral responsibility is "plainly useless."

[8] Each author in *ibid* identifies cases of this form.

[9] For instance, Pettit (2007, p. 175) discusses autonomous agency in a choice context ("value relevance"), the ability to make judgments about choices (including access to relevant evidence) ("value judgment"), and control over a choice ("value sensitivity"). Braham and Van Hees (2011, p. 7) discuss an "agency condition" requiring the subject be "capable of planning and forming intentions …[and] distinguishing right from wrong and good from bad"; a "causal relevancy condition" requiring "a causal relation between the action of the agent and the resultant state of affairs"; and an "avoidance opportunity condition" requiring the agent have "a reasonable opportunity to do otherwise." Similar conditions on responsibility feature in many mainstream accounts, including Shoemaker's (2011, 2015).

[10] See also Collins (2017b, p. 59).

[11] Duijf (2018, p. 435) offers conditions under which voids can exist but does not provide formal criteria, instead largely appealing to lacks in appropriate loci of responsibility.

[12] See also note [2].

[13] This response may also be available to the monist who suggests that accountability and answerability form a single phenomenon with different practices filling apparent deficiencies in one element of that phenomenon. However, I am not aware of scholarship taking this tack and so introduce this as a pluralism-based complication.

[14] de Sio and Mecacci (2021, pp. 1073–1074) describe this as a form of "deflationism."

[15] See, e.g., Copp (2006, p. 219n52). *Cf*. Kiener (2022).

REFERENCES

Bovens, M. (1998). *The Quest for Responsibility*. Cambridge UP.
Braham, M., & van Hees, M. (2011). Responsibility Voids. *The Philosophical Quarterly*, *61*(242), 6–15.

Buell, S. W. (2018). The Responsibility Gap in Corporate Crime. *Criminal Law and Philosophy*, *12*(3), 471–491. https://doi.org/10.1007/s11572-017-9434-9

Champagne, M., & Tonkens, R. (2015). Bridging the Responsibility Gap in Automated Warfare. *Philosophy and Technology*, *28*(1), 125–137. https://doi.org/10.1007/s13347-013-0138-3

Chapman, B. (1998). More Easily Done than Said. *Oxford Journal of Legal Studies*, *18*, 293–329.

Collins, S. (2017a). Filling Collective Duty Gaps. *Journal of Philosophy*, *114*(11), 573–591. https://doi.org/10.5840/jphil20171141141

Collins, S. (2017b). Duties of Group Agents and Group Members. *Journal of Social Philosophy*, *48*(1), 38–57. https://doi.org/10.1111/josp.12181

Collins, S. (2019). Collective Responsibility Gaps. *Journal of Business Ethics*, *154*(4), 943–954. https://doi.org/10.1007/s10551-018-3890-6

Copp, D. (2006). On the Agency of Certain Collective Entities. *Midwest Studies in Philosophy*, *30*(1), 194–221. https://doi.org/10.1111/j.1475-4975.2006.00135.x

Danaher, J. (2016). Robots, Law and the Retribution Gap. *Ethics and Information Technology*, *18*(4), 299–309. https://doi.org/10.1007/s10676-016-9403-3

Danaher, J. (2022). Tragic Choices and the Virtue of Techno-Responsibility Gaps. *Philosophy and Technology*, *35*(2), 26. https://doi.org/10.1007/s13347-022-00519-1

de Jong, R. (2020). The Retribution-Gap and Responsibility-Loci Related to Robots and Automated Technologies. *Science and Engineering Ethics*, *26*(2), 727–735. https://doi.org/10.1007/s11948-019-00120-4

de Sio, F. S., & Mecacci, G. (2021). Four Responsibility Gaps with Artificial Intelligence. *Philosophy and Technology*, *34*(4), 1057–1084.

Duijf, H. (2018). Responsibility Voids and Cooperation. *Philosophy of the Social Sciences*, *48*(4), 434–460. https://doi.org/10.1177/0048393118767084

Enoch, D. (2012). Being Responsible, Taking Responsibility, and Penumbral Agency. In U. Heuer & G. Lang (Eds.), *Luck, Value, and Commitment* (pp. 95–132). Oxford UP.

Glavaničová, D., & Pascucci, M. (2022). Vicarious Liability. *Ethics and Information Technology*, *24*(3), 28. https://doi.org/10.1007/s10676-022-09657-8

Grübler, G. (2011). Beyond the Responsibility Gap. *AI & Society*, *26*(4), 377–382. https://doi.org/10.1007/s00146-011-0321-y

Himmelreich, J. (2019). Responsibility for Killer Robots. *Ethical Theory & Moral Practice*, *22*(3), 731–747. https://doi.org/10.1007/s10677-019-10007-9

Hindriks, F., & Veluwenkamp, H. (2023). The Risks of Autonomous Machines: From Responsibility Gaps to Control Gaps. *Synthese*, *201*(1), 21. https://doi.org/10.1007/s11229-022-04001-5

Kiener, M. (2022). Can We Bridge AI's Responsibility Gap at Will? *Ethical Theory & Moral Practice*, *25*(4), 575–593. https://doi.org/10.1007/s10677-022-10313-9

Köhler, S., Roughley, N., & Sauer, H. (2018). Technologically Blurred Accountability? In C. Ulbert, P. Finkenbusch, E. Sondermann, & T. Debiel (Eds.), *Moral Agency and the Politics of Responsibility*. Routledge.

Königs, P. (2022). Artificial Intelligence and Responsibility Gaps. *Ethics and Information Technology*, *24*(3), 36. https://doi.org/10.1007/s10676-022-09643-0

Kornhauser, L. G., & Sager, L. A. (1993). The One and the Many. *California L.R.*, *81*(1), 1–59. https://doi.org/10.2307/3480783

Lawford-Smith, H., & Collins, S. (2017). Responsibility for States' Actions. *Philosophy Compass*, *12*(11), e12456. https://doi.org/10.1111/phc3.12456

List, C. (2006). The Discursive Dilemma and Public Reason. *Ethics*, *116*(2), 362–402. https://doi.org/10.1086/498466

List, C. (2021). Group Agency and Artificial Intelligence. *Philosophy and Technology*, *34*(4), 1213–1242. https://doi.org/10.1007/s13347-021-00454-7

List, C., & Pettit, P. (2011). *Group Agency*. Oxford UP.

Matthias, A. (2004). The Responsibility Gap. *Ethics and Information Technology*, *6*(3), 175–183. https://doi.org/10.1007/s10676-004-3422-1

Moen, L. J. K. (2024). Against Corporate Responsibility. *Journal of Social Philosophy*, *55*(1), 44–61. https://doi.org/10.1111/josp.12547

Mukerji, N., & Luetge, C. (2014). Responsibility, Order Ethics, and Group Agency. *Archiv für Rechts- Und Sozialphilosophie*, *100*(2), 176–186. https://doi.org/10.25162/arsp-2014-0013

Nyholm, S. (2017). Attributing Agency to Automated Systems. *Science and Engineering Ethics*, *24*, 1–19.

Oimann, A.-K. (2023). The Responsibility Gap and LAWS. *Philosophy and Technology*, *26*, 3.

Pettit, P. (2001). Deliberative Democracy and the Discursive Dilemma. *Philosophical Issues*, *11*(1), 268–299. https://doi.org/10.1111/j.1758-2237.2001.tb00047.x

Pettit, P. (2007). Responsibility Incorporated. *Ethics*, *117*(2), 171–201. https://doi.org/10.1086/510695
Shoemaker, D. (2011). Attributability, Answerability, and Accountability. *Ethics*, *121*(3), 602–632. https://doi.org/10.1086/659003
Shoemaker, D. (2015). *Responsibility from the Margins*. Oxford UP.
Smith, A. M. (2012). Attributability, Answerability, and Accountability. *Ethics*, *122*(3), 575–589. https://doi.org/10.1086/664752
Smith, T. (2009). Non-Distributive Blameworthiness. *Proceedings of the Aristotelian Society*, *109*(1), 31–60. https://doi.org/10.1111/j.1467-9264.2009.00257.x
Sondermann, E., Ulbert, C., & Finkenbusch, P. (2018). Introduction. In C. Ulbert, P. Finkenbusch, E. Sondermann, & T. Debiel (Eds.), *Moral Agency and the Politics of Responsibility* (pp. 1–18). Routledge.
Sparrow, R. (2007). Killer Robots. *Journal of Applied Philosophy*, *24*(1), 62–77. https://doi.org/10.1111/j.1468-5930.2007.00346.x
Swoboda, T. (2017). Autonomous Weapon Systems. In V. C. Müller (Ed.), *Philosophy and Theory of Artificial Intelligence* (pp. 302–313). Springer.
Thompson, D. F. (1980). The Quest for Responsibility. *American Political Science Review*, *74*(4), 905–916. https://doi.org/10.2307/1954312
Tigard, D. W. (2021). There Is No Techno-Responsibility Gap. *Philosophy and Technology*, *34*(3), 589–607. https://doi.org/10.1007/s13347-020-00414-7
Tollon, F. (2022). Responsibility Gaps and the Reactive Attitudes. *AI and Ethics*, *3*(1), 295–302. https://doi.org/10.1007/s43681-022-00172-6
van de Poel, I., Nihlén Fahlquist, J., Doorn, N., Zwart, S., & Royakkers, L. (2012). The Problem of Many Hands. *Science and Engineering Ethics*, *18*(1), 49–67. https://doi.org/10.1007/s11948-011-9276-0
Zając, M. (2020). Punishing Robots. *Journal of Military Ethics*, *19*(4), 285–291. https://doi.org/10.1080/15027570.2020.1865455

AUTHOR BIOGRAPHY

Michael Da Silva is an Associate Professor at the University of Southampton in Southampton, United Kingdom. His published works include articles in the Journal of Social Philosophy, Ethical Theory and Moral Practice, Bioethics, and the European Journal of Political Theory.

How to cite this article: Da Silva, M. (2024). Responsibility gaps. *Philosophy Compass*, e70002. https://doi.org/10.1111/phc3.70002