COMMENTARY

Workplace automation without achievement gaps: a reply to Danaher and Nyholm

Daniel W. Tigard[1] ⃝

Abstract

In a recent article in this journal, John Danaher and Sven Nyholm raise well-founded concerns that the advances in AI-based automation will threaten the values of meaningful work. In particular, they present a strong case for thinking that automation will undermine our achievements, thereby rendering our work less meaningful. It is also claimed that the threat to achievements in the workplace will open up 'achievement gaps'—the flipside of the 'responsibility gaps' now commonly discussed in technology ethics. This claim, however, is far less worrisome than the general concerns for widespread automation, namely because it rests on several conceptual ambiguities. With this paper, I argue that although the threat to achievements in the workplace is problematic and calls for policy responses of the sort Danaher and Nyholm outline, when framed in terms of responsibility, there are no 'achievement gaps'.

Keywords Automation · Work · Achievement · Responsibility · Technology ethics · AI ethics · Robotics

1 Introduction

The development of increasingly sophisticated technology in the workplace has long been a source of social and political upheaval. As some authors have recently suggested, we can look to the Luddites of the nineteenth century industrial revolution, if not farther back, to see how workplace automation might promote the well-being of some—particularly the owners of automated workforces—while hindering others' well-being [1, 2]. Considering the contemporary rise in AI and robotics research, and the growing prevalence of such systems throughout our personal and professional lives, it is understandable and important that many of these familiar discussions are being raised anew.

In a recent article in this journal, John Danaher and Sven Nyholm present a strong case for thinking that AI-based automation is posing a newfound threat to the values of meaningful work [3].[1] Specifically, human achievement is aptly framed as a valuable feature of work, one that may indeed trouble us to lose sight of. Danaher and Nyholm

argue that by automating work, the increasing use of AI and robotic technology stands to undermine the possibility of human achievement, thereby rendering work less meaningful. They also claim that this threat will open up 'achievement gaps'—a term meant to signify the flipside of the 'responsibility gap', which is now commonly discussed in the ethics of technology, particularly concerning AI and robotics.

Responsibility gaps are almost unanimously seen as a highly undesirable result of automating technologies, since, by definition, they entail troubling circumstances where no one is clearly responsible. Accordingly, if in our increasing use of AI and robotic systems we find scenarios that resemble responsibility gaps, we may well have reason to be alarmed. Fortunately, when framed in terms of the responsibility gap, it appears there are no such gaps in workplace achievement. This is not to say that the threat to achievements from automation should not concern us. Like the general concerns for widespread automation, the threat to achievements in the workplace is problematic and, I believe, calls for policy responses of the sort Danaher and Nyholm outline. However, as I will argue, we need not worry that workplace automation will bring about novel instantiations

✉ Daniel W. Tigard
daniel.tigard@tum.de

1 Institute for History and Ethics of Medicine, Technical University of Munich, Ismaninger Str. 22, 81675 Munich, Germany

---

[1] Unless otherwise specified, when referencing Danaher and Nyholm, I will refer to this work [3]. Page numbers of the Online First version will follow any direct quotes.

of the responsibility gap. To show this, I will first revisit the notions of meaningful work, achievement, and the threat from automation, as presented by Danaher and Nyholm—and I largely accept them on these terms. Next, I clarify the nature of responsibility and point to a mismatch in Danaher and Nyholm's analogy with achievement. Lastly, I show how the threat to achievements in the workplace should not generate worries akin to the responsibility gap and I consider some potential objections. Again, my clarification should come as a reassurance against newfound worries of a responsibility gap, and so, I close on a relatively optimistic note.

2 Meaningful work, achievement, and the threat from automation

Danaher and Nyholm begin by establishing a purely economic notion of work, that is, any activity that 'consists of skills (physical, cognitive, and emotional) that are performed by individuals in return for some kind of economic reward' (p. 2). Following Danaher's fuller treatment of the ethics of automating technologies [1], this notion is deliberately narrow so as to exclude unpaid activities like hobbies and caregiving, which some people may be inclined to think of as work. However, these latter activities raise a host of additional considerations, including the thought that we arguably cannot or should not outsource them to technology. Accordingly, by narrowing the scope, we can set aside some of the additional puzzles and clearly articulate the values associated with work in the purely economic sense.[2]

Not all work is meaningful work, of course. To make this distinction, Danaher and Nyholm rely on Susan Wolf [5], namely Wolf's account of what makes life in general meaningful. In a subjective sense, one must be 'engaged in activities and projects that she is passionate about, and that can also be recognized as valuable from a wider, not purely subjective perspective' (p. 3). This twofold subjective–objective framing likewise applies to our evaluations of work. Here we find support for what is likely a common intuition, that is, that some lines of work are more meaningful than others. For example, if one were to count blades of grass all day in exchange for monetary compensation, it would be difficult to say this activity is somehow valuable from the more objective viewpoint, even if it brought the person a great sense of subjective fulfillment.[3] By contrast,

one's work might be greatly valued from the wider point of view but fail to give the subject any personal sense of satisfaction. In either case, we see reason to question the extent to which the person's work is truly meaningful. With this in mind, Danaher and Nyholm characterize meaningful work with two conditions: value in 'the overarching output produced by the workplace', and value in 'the individual worker's job' and sub-tasks (p. 3).

Considering this characterization of meaningful work, we begin to see how our engagement in this sort of activity carries the potential for bringing about a sense of human achievement. What, then, is achievement? Danaher and Nyholm develop a rich composite account by combining key aspects of Gwen Bradford's work [8] and that of Hannah Maslen, Julian Savulescu and Carin Hunt [9]. To briefly summarize, it is said that 'whenever we are assessing the value of someone's achievements, the following four variables need to be kept in mind': (1) the value of the output produced, especially objective value; (2) the nature of the causal contribution of the agent, i.e. a non-lucky result; (3) the cost of the agent's commitment, e.g., more time or effort typically indicates a greater achievement; and (4) the voluntariness of the agent's actions (p. 5). Again, where any one of these features becomes threatened or where any one is lacking entirely, we would have reason to question whether or not the agent's activity has truly led to an achievement. As Danaher and Nyholm argue, this is precisely what we see with the emergence of automating technologies in the workplace.

To see how automation threatens human achievement at work, Danaher and Nyholm invite us to consider two different forms of workplace automation. On one hand, there is the 'total replacement' of humans, where machines render human workforces redundant. Here, it seems clear that the possibility of human achievement is entirely removed, simply because the humans—no matter how skilled—'no longer have access to any form of workplace achievement' (p. 6). On the other hand, there are 'collaborative displacements', where human workers are redeployed such that they must 'collaborate with the machines to produce the output'; this typically leads to 'a redrawing of the boundaries' of a job, with humans supervising machines, humans working merely to maintain machines, or machines supervising humans (p. 3). In these sorts of cases, it becomes less obvious that the potential for human achievement is undermined. Nonetheless, as Danaher and Nyholm show, three of the four key aspects of achievements are still threatened by collaborative displacements resulting from automation.

First, automation threatens achievements in the workplace by reducing the value of the outputs. This can be seen by considering cases where humans merely maintain the machines' functioning, also where humans take orders from machines, and even where humans supervise machines.

---

[2] Here, I should flag that my purposes throughout this section are limited to sketching the notions of meaningful work, achievement, and the threat from automation. Readers might take issue with some of this framework, and I encourage them to examine Danaher and Nyholm's work on such topics [1, 4].

[3] This scenario has, however, provided widespread value to the philosophical community, notably Rawls [6], Setiya [7], and others.

While in these latter scenarios, humans 'retain creative control and mastery' (p. 7), Danaher and Nyholm note that such positions of power are often held by an elite few, thereby depriving most workers of opportunities for achievements; and with the use of machine-learning systems, even the elite become removed from the outputs.[4] Second, they argue that 'automation, almost by necessity, reduces the cost of the human commitment' (p. 7), namely because such technologies are developed and deployed for this exact reason. That is, we aim at reducing the time, effort, or stress we once had to exert at work. What we thereby decrease, however, is a sense of achievement in the outputs. Third, automation reduces the causal contribution of the agent. For Danaher and Nyholm, this is 'again, almost by necessity' (p. 8), since again we create and deploy such technologies precisely because we want to be able to contribute less while retaining or perhaps increasing the outputs, even if we then have a weaker claim to those outputs constituting an achievement. Fourth, they note that automation might not present a marked decrease in the voluntariness of work; however, the threat to the first three features is enough, on their account, to show that automating technologies have 'the potential to open up numerous achievement gaps' (p. 8).

3 Achievement and responsibility

Do achievement gaps really constitute 'responsibility gaps'? To be sure, various lines of counterargument could be raised with respect to the notions covered so far, and indeed, Danaher and Nyholm foreshadow and respond to several contrasting positions on the ideas of meaningful work, achievement, and the threat from automation. But as clarified at the outset, I accept them on these terms. For this reason, it seems to me that the policy responses they outline will be important to consider in our efforts to ensure our continued well-being in light of increasingly prevalent automation. Still, where I think we can resist any additional cause for concern is in their framing of the achievement gap as a gap in *responsibility*. Accordingly, to show how we can resist the added concerns associated with responsibility gaps, I must first clarify several crucial features of responsibility, more generally.

First, it is commonly acknowledged in the philosophical literature that responsibility must somehow "attach" to agency. For instance, theorists like Michael McKenna [10] articulate the intuition that morally *responsible* agency is

usually thought of as a subset of moral *agency*, a concept often framed in terms of consciousness, autonomous action, responsiveness to reasons, and so on.[5] The subset relationship makes sense conceptually, but also in practice. Consider that if someone is being held morally responsible (often via praise or blame), she must first be a moral agent. If someone or something is not clearly a moral agent—consider trees, dogs, fetuses—it would be similarly unclear how we could sensibly hold it responsible.

Next, as I have just now suggested, at least implicitly, responsibility is something that *happens*, namely when we make evaluations of others (or of ourselves). It may be a course of action, a perceived attitude, or perhaps the underlying character of the agents we are evaluating. The basic point to be made here is that responsibility is a dynamic, often interactive process that takes place in a world where moral agents deliberate, decide, take action (and so on), and where such events elicit the *responses* of those who are in some way affected.[6] Granted, some might challenge this picture, maintaining instead that responsibility is a property or quality, something we see in the world, perhaps in individuals with various obligations or well-mannered traits. In these ways, we naturally speak of *responsible* individuals.[7] Undoubtedly, a full argument for either view of responsibility cannot be given here, and fortunately, I need not provide one. Although they do not explicitly acknowledge their position, Danaher and Nyholm lend support to the more interactive view, saying, for instance, 'we blame ourselves and others for doing bad things, we also praise ourselves for achieving positive (or value neutral) things' (p. 9). Accordingly, I will adopt and expand upon this highly fruitful framing of responsibility.

The more relational, interactive view can be illustrated with a concise formula set out by David Shoemaker: 'To be a responsible agent is to be worthy of X for Y in virtue of Z' [15, p. 17]. Here, X represents the variety of *responses* on behalf of those affected, where the positive responses are usually generalized into "praise" and the negative are generally related to "blame". As Shoemaker explains, 'Y refers to something like actions or attitudes' [15]. We praise or blame or otherwise evaluate others (including ourselves) in light of something that they are *doing* or something they did, or

---

[4] As a great deal of recent literature attests, the use of "black box" machine-learning systems may undermine human understanding of how an output is given, even among those who design, program, or use such systems. Accordingly, it becomes difficult to attribute those outputs to any human. That is, while the results might be valuable, they are not properly the *achievements* of any human.

[5] Of course, I'm glossing over countless questions addressed throughout volumes of work on moral agency. However, the exact characterization is largely inconsequential here.

[6] Indeed, a growing body of work in theoretical and applied ethics supports the general framing of responsibility as dependent upon a great variety of relationships and manners of responding to one another (and to ourselves). Among others, see [11–13].

[7] See my recent paper in this journal [14], for distinctions between various uses of 'responsibility', particularly in applied settings (e.g. responsible innovation).

because of some *way* that they did it. Lastly, Z represents what Shoemaker calls the 'responsibility-maker'—that is, the capacity of the agent that renders our response appropriate. For example, I might blame a friend for forgetting my birthday, because she *should have known*. Here, the fact that my friend *should have known*—pointing to an epistemic capacity—renders my blame a fitting response to her forgetting my birthday.[8]

With the responsibility formula in mind, a preliminary ambiguity can be seen in the account of Danaher and Nyholm. To begin their analysis of achievement, they claim that 'achievements are, in essence, a positive manifestation of responsibility' (p. 4). They also accept that responsibility is a process of responding to one another and to ourselves, namely with blame for bad things and praise for positive (or value neutral) things. That is, blaming and praising (and other related evaluations, both negative and positive) *are the responses* that make up our responsibility practices—i.e., the "X" factor in the above formula. Danaher and Nyholm very reasonably take achievement to be the sort of thing that often plays a role in our evaluations. However, an achievement on its own is surely not a responsibility *response*. Indeed, it sounds quite odd to say an achievement alone—absent anything resembling praise or blame—is a way of holding someone responsible. Instead, it is the kind of thing we hold others or ourselves responsible *for*—i.e., the "Y" factor.

Before turning to the possible gaps in responsibility, we should take careful note of the mismatch in variables. We blame for bad things; we praise for good or neutral things. Danaher and Nyholm set out to 'draw explicit analogies' between achievement gaps and responsibility gaps, but achievements are merely among the *things*. They are not *responses* to things. As such, relative to our responses, achievements play only a peripheral, instrumental role in the process of holding one another responsible. Since it seems clear that achievements are the "Y" factor, to say there are 'achievement gaps' would be much like saying there are gaps in honesty, kindness, life saving, and so on—that is, the actions and attitudes that we plausibly hold one another (and ourselves) responsible *for*. Granted, a more charitable reading of achievement gaps can be discerned, but as I show next, this prospect should not worry us.

4 Easing the worries over gaps

In this section, I turn to an analysis of the 'responsibility gap' supposedly brought about by AI and robotic technologies. Building upon the mismatch between achievement and responsibility identified above, here I make clear why the threat to achievements in the workplace (perhaps also elsewhere) should not generate worries akin to the responsibility gap.

Much of the recent work on the responsibility gap draws its inspiration from Andreas Matthias's essay of that title [17]. As Matthias was concerned to show, machines today are reaching a level of sophistication such that they will be able to act in ways that cannot be traced back to their manufacturer or their operator. This is because neither the manufacturer nor the operator will have sufficient knowledge or control of the machine's actions. As is widely accepted, knowledge and control are key components for appropriately holding someone responsible—i.e., examples of "Z" factors in the formula above.

But what exactly is the gap? In a helpful essay, Sebastian Köhler, Neil Roughley, and Hanno Sauer [18] frame the responsibility gap as a 'normative mismatch' in the sense that our usual theories of moral (and legal) responsibility should be equipped to identify the responsible parties. And yet, due to the possibility of machine-learning systems acting on their own and for reasons that remain opaque to users and designers, we are often left unable to properly locate responsibility.[9] In short, responsibility gaps are seen when the following two conditions obtain: (1) it seems fitting to hold someone responsible for some Y; but (2) there is no candidate who it is fitting to hold responsible for Y.[10] With a brief consideration of each condition, it is clear that both are necessary for gaps to arise. The first is simply one's observation of an event—an action, attitude, etc.—and corresponding inclination to respond to the source. Without the observation, event, or inclination to respond, there would be no process of responsibility initiated in the first place. The second must also hold, for this describes the absence of any potential source. If one were present, there would be no gap in responsibility.[11]

[8] This summary is a highly generalized depiction of an elaborate framework for illustrating responsibility. But again, a more detailed review is beyond my purposes, and the basics should suffice. For more on responsibility and the ignorance excuse, see Biebel [16].

[9] I should note that Köhler, Roughley, and Sauer argue against the presence of gaps, but narrowly in terms of accountability only. For even wider resistance to the idea of a technology-based responsibility gap, see [19].

[10] See Köhler, Roughley, and Sauer [18, p. 54]. Here they include a qualification concerning the *degree* to which we might hold someone responsible—and while helpful, I leave it aside for sake of brevity. I've also used "Y" to stay consistent with the formula above, and expanded from accountability to responsibility generally.

[11] Another complexity that cannot be fully addressed here is the idea that, even where the source *is* present, they might still deny any ill-will, e.g. by offering an excuse or justification. Thus, it may seem

The questions for Danaher and Nyholm then become: what exactly does the gap in *achievement* look like, and where (if anywhere) would it occur? To address the latter inquiry, it appears fruitful again to take Danaher and Nyholm on their terms, and so, I will analyze each form of automation they establish. But first, I will need to reframe the responsibility gap in terms of workplace achievement, and it seems that the most charitable account would be the following: (1) it seems fitting to praise someone for some workplace achievement; but (2) there is no candidate who it is fitting to praise for that workplace achievement.[12] Again, both conditions are necessary in order to truly say there is a gap, namely in achievement. Yet, as I will show, we cannot affirm both conditions on any form of automation Danaher and Nyholm establish.

First, there is 'total replacement' of humans by automating technologies. Danaher and Nyholm quickly assessed this scenario as clearly putting 'an end' to workplace achievement, since the humans who are made redundant 'no longer have access to any form of workplace achievement' (p. 6). In terms of the achievement gap, here we can affirm that (2) there is no candidate to praise. However, it would be far from clear that (1) it seems fitting to praise someone. If we are considering automation scenarios where human workers are entirely replaced, there simply is no one we might be inclined to praise in the first place. Furthermore, on the total replacement scenario, we would have nothing to praise anyone *for*, namely since the supposed achievements at stake are not properly seen as achievements. Indeed, these were characterized in noticeably human terms: a costly commitment of the agent, voluntariness, and so on. Thus, although the total replacement form of automation entails the elimination of human achievement in the workplace, on this scenario, there is also no achievement gap.

Next, there are the three forms of 'collaborative displacement'. I will begin with the simplest case and progress to the more complex. Recall that some forms of collaborative displacement entail humans working to maintain machines. On this scenario, it seems relatively straightforward that there is no achievement gap, for similar reasons seen in

cases of total replacement. As Danaher and Nyholm explain, the maintenance work by humans may be quite sporadic— and even when needed, humans merely 'step in to repair or fix the machines, or reprogram/repurpose them' (p. 3). That is, on the whole, human involvement is not necessary to produce the output, much like with total replacement. Consider also that where a human must temporarily step in, we may be inclined to praise them for the continuation, or perhaps renewal, of the flow of workplace achievements. In other words, we might affirm condition (1); however, in looking to the human maintenance workers, we find a candidate to praise and thereby cannot affirm (2) that there is no candidate.

Then, there are the collaborative displacements where machines supervise or direct human workers. On this form of collaboration, I suggest, it will be difficult to affirm condition (1). Danaher and Nyholm describe cases wherein humans follow orders of machine-learning systems, and it is the machine that does any creative or intellectual work. For example, an algorithm can use a database 'to figure out the best way to make a fuel efficient, aerodynamic car…[then] humans go off and build the car' (p. 3). Here, it appears that although some may be inclined to praise the humans who follow the machines' orders, it seems unlikely that we would be inclined to praise them *for the workplace achievement*—in this case, the excellently designed car. Instead, we might praise them for their careful attention to pre-existing design details, for instance, or perhaps for their flexibility and humility in being redeployed in ways that leave them subservient to artificially intelligent systems. Also, notice again that even where we are inclined to praise the human order followers, and where we do indeed have the workplace *achievement* in mind, we thereby find a candidate for our praise. In this way, the fulfillment of condition (1) again leaves us unable to affirm (2), so there is still no gap.

Lastly, consider the scenario where humans supervise machines. Here, it may be most tempting to think that automation brings about an achievement gap. After all, in these sorts of cases, human workers are redeployed but are still present and in-charge, and thereby should be able to claim the fruits of their labor, including any praise-like responses. But do the humans who retain supervisory roles in cases of human–machine collaboration really deserve praise for the workplace achievements? Consider a human supervisor at a household goods distribution center.[13] This person is in command of all local logistics and may even retain some 'creative control and mastery', as Danaher and Nyholm put

---

Footnote 11 (continued)

that a gap has reemerged; yet, the responsibility *response* has still found its target (even if it becomes inappropriate to continue holding that person responsible). On McKenna's conversational model, our response still initiates the responsibility "exchange.".

[12] As an anonymous reviewer aptly suggests, when it comes to workplace achievement, there is often no systematic incentive to identify anyone to praise; whereas, in cases of damages, there are likely firm mechanisms in place—such as legal systems to help compensate victims. I take this as a further challenge for Danaher and Nyholm, considering their aim to 'draw explicit analogies' between the two sorts of cases.

---

[13] Amazon likely comes to mind for many, and I resist considering Jeff Bezos in particular—for he may be low-hanging fruit for my argument, since readers might agree that it does not seem fitting to praise such supervisors.

it. Nonetheless, here, we again run into the same dissolution of the achievement gap seen in other forms of collaborative displacement. That is, *if* we look to the human in-charge and affirm that (1) it seems fitting to praise that person for the workplace achievement, we thereby affirm that there is a candidate who it is fitting to praise. Without affirming condition (2), then, we see that there is no achievement gap. Moreover, to many, it will seem far-fetched to look to the humans who supervise machines and still be inclined to praise in the first place. As Danaher and Nyholm show, for the few elite humans who retain creative or supervisory workplace roles, we see a 'loss in the value of work-related outputs'; we see a reduction in the cost of their commitment (i.e., less time, effort, and stress is exerted); and their causal connection to the output is typically severed (pp. 7–8). Hence, for many, it will be difficult to affirm that (1) it seems fitting to praise someone for the workplace achievement. Accordingly, it appears there is no gap.

Before concluding, I should consider a few potential lines of objection, one of which arises from what Danaher calls the 'retribution gap' [20]. On Danaher's account, much like the one I have established here, the gap is framed as a mismatch. Specifically, retribution gaps occur when we have mismatches between 'the human desire for retribution and the absence of appropriate subjects of retributive blame' [20, p. 209]. This framing could be used to formulate a plausible rendition of achievement gaps—namely, mismatches between the human desire to praise and the absence of appropriate subjects of praise. One could then say there are gaps by pointing to cases where there are no appropriate subjects, yet there is still a strong psychological desire to praise someone. In response, consider again the nature of gaps in responsibility generally, whether negative or positive. It seems that a truly bothersome gap-like situation will be one wherein there are no appropriate targets of our responses, but we nonetheless find it fitting to blame or praise, and so on. In contrast, the situations Danaher describes involve a desire, not a judgment of fittingness. No doubt, there will be lots of cases where our desires to mete out blame or praise go unfulfilled. But this speaks more to our psychological make-up, and likely an excessive tendency to seek out responsibility, than it does to the coherence of our moral practices. In other words, a desire to hold someone responsible does not entail that it seems fitting to do so. While we might often wish for someone to blame (or praise) when things go poorly (or well), this desire is not enough to affirm the first condition of responsibility gaps.[14]

As a second line of objection, it might be thought, particularly on the human-supervisor form of automation, that we do see a possible achievement gap. Specifically, one might be inclined to praise the humans who retain creative control or mastery over a highly efficient automated workforce—i.e., one might affirm (1) that it seems fitting to praise. Yet, imagine that sometime later, the potentially praise-giving person learns more about the workplace conditions, namely that the output is largely automated, the humans in-charge contribute little time or effort, and that the output is far downstream from any human involvement. In this way, it may well be that (2) there is no candidate—including the human supervisors—who it is fitting to praise. To be sure, a similar temporal progression might be seen in response to other forms of collaborative displacement, revealing more potential achievement gaps in cases of human–machine collaborations.

This line of thought is worth taking seriously, but notice that the objection rests upon taking apart the two necessary conditions. That is, here, we imagined that one *initially* finds it fitting to praise someone, but also that *later*, they find there is no fitting candidate. However, achievement gaps are supposed to be closely related to responsibility gaps, and for this reason, one must be able to simultaneously affirm both necessary conditions. Why is this? Imagine, for example, that a long-lost friend sends me a birthday greeting. I did not expect them to remember, and I am inclined to praise them (express gratitude, etc.) for their surprising generosity. Then, I learn that the greeting was an automated message sent from their digital calendar, requiring no awareness on behalf of the friend and certainly no generosity. It is safe to suppose here that my praise-like response will be modified in light of this new information. I may realize that (2) there is no candidate to praise, but notice that a modification also takes place in my initial inclination to praise: I come to realize that it is in fact not fitting—i.e. condition (1) is no longer affirmed.

It seems that our responses would undergo a similar modification in cases of human supervisors in automated workplaces, and indeed on each form of human–machine collaboration established by Danaher and Nyholm. Once we affirm that there is no candidate to praise, we come to realize that it is not fitting to be inclined to praise anyone in the first place. Thus, there is no mismatch between our inclinations to praise anyone for workplace achievements and the presence of a candidate to praise. In other words, there are no 'gaps' in achievement.

5 Conclusion

In closing, it is worth stopping to ask: Who exactly is the primary subject of "harm" (broadly speaking) in the supposed gap scenarios? Typically, in cases of responsibility

---

[14] If 'gaps' occurred every time we *desire* holding someone responsible, it would appear that gaps are extremely prevalent and perhaps should not be taken so seriously. Instead, it seems that a more grounded moral judgment is needed, which is why I rely on the notion of fittingness. For a helpful overview, see [21].

gaps, the harm is seen as falling upon the person inclined to respond (usually with blame) and finding no one to respond to. This is often because they seek apologies or some sort of remuneration, and as we can imagine, it sets back their interests when such demands remain unfulfilled. But what about cases of achievement gaps? If we want to draw truly close analogies between the two scenarios, we would consider the subject of harm to be the person inclined to respond with praise and finding no one to praise. And perhaps there is some degree of disappointment here, but it hardly seems to be a worrisome kind of experience for that person. With this in mind, we might say there is yet another mismatch between responsibility gaps and achievement gaps. Nevertheless, on the account of Danaher and Nyholm, the harm is seen as falling upon the humans who miss out on achieving something in the workplace. But on that picture, we run into a sort of non-identity problem—for as soon as we identify the subjects of this kind of harm, we thereby affirm that it is not fitting to praise them for the workplace achievement, and so they cannot really be harmed in this way.

To be clear, workplace automation undoubtedly raises a host of challenges, including the potential for missed opportunities to achieve something in the workplace. Accordingly, Danaher and Nyholm aptly suggest policy responses that include emphasizing other aspects of meaningful work, finding ways to retain a 'human touch' on the final outputs, placing greater emphasis on teamwork, and finding other means of fulfilling (non-work) achievements (pp. 8–9). Policies to ensure our continued well-being are quite reasonable considering the growing prevalence of automation and possibility of losing sight of human achievement in the workplace. However, these efforts need not help us to remedy any sort of gap in responsibility, for once we fully consider what it means to face such scenarios, we see that they do not really come about. This should come as a word of comfort to those who worry that AI and robotic technologies will generate newfound instantiations of the responsibility gap. When framed in terms of responsibility, it appears there are no achievement gaps.

Declarations

References

1. Danaher, J.: Automation and Utopia: human flourishing in a world without work. Harvard University Press (2019)
2. Smith, D.: Perhaps Ned Ludd had a point? AI and Society, 1–3 (2021)
3. Danaher, J., Nyholm, S.: Automation, work and the achievement gap. AI and Ethics, 1–11 (2020)
4. Nyholm, S.: Humans and robots: ethics, agency, and anthropomorphism. Rowman and Littlefield (2020)
5. Wolf, S.: Meaning in life and why it matters. Princeton University Press (2010)
6. Rawls, J.: A theory of justice. Belknap Press (1971)
7. Setiya, K.: Cognitivism about instrumental reason. Ethics **117**(4), 649–673 (2007)
8. Bradford, G.: Achievement. Oxford University Press (2016)
9. Maslen, H., Savulescu, J., Hunt, C.: Praiseworthiness and motivational enhancement: 'no pain, no praise'? Aust J Philos **98**(2), 304–318 (2020)
10. McKenna, M.: Conversation and responsibility. Oxford University Press (2012)
11. Strawson, P.F.: Freedom and resentment. Proc. Br. Academy **48**, 1–25 (1962)
12. Wallace, R.J.: Responsibility and the moral sentiments. Harvard University Press (1994)
13. Coeckelbergh, M.: Robot rights? Towards a social-relational justification of moral consideration. Ethics Inf Technol **12**(3), 209–221 (2010)
14. Tigard, D.W.: Responsible AI and moral responsibility: a common appreciation. AI and Ethics, 1–5 (2020)
15. Shoemaker, D.: Responsibility from the margins. Oxford University Press (2015)
16. Biebel, N.: Epistemic justification and the ignorance excuse. Philos. Stud. **175**(12), 3005–3028 (2018)
17. Matthias, A.: The responsibility gap: ascribing responsibility for actions of learning automata. Ethics Inf. Technol. **6**(3), 175–183 (2004)
18. Köhler, S., Roughley, N., Sauer, H.: Technologically blurred accountability? In: Ulbert, C., et al. (eds.) Moral agency and the politics of responsibility. Routledge (2017)
19. Tigard, D.W.: There is no techno-responsibility gap. Philos. Technol. 1–19 (2020)
20. Danaher, J.: Robots, law and the retribution gap. Ethics Inf. Technol. **18**(4), 299–309 (2016)
21. Jacobson, D.: Fitting attitude theories of value. Stanford Encyclopedia of Philosophy, E. Zalta (ed.), https://plato.stanford.edu/archives/spr2011/entries/fitting-attitude-theories/ (2011)