

Automation, Work and the Achievement Gap

By John Danaher and Sven Nyholm

Forthcoming in *AI and Ethics*

Abstract: Rapid advances in AI-based automation have led to a number of existential and economic concerns. In particular, as automating technologies develop enhanced competency they seem to threaten the values associated with meaningful work. In this article, we focus on one such value: the value of achievement. We argue that achievement is a key part of what makes work meaningful and that advances in AI and automation give rise to a number of achievement gaps in the workplace. This could limit people's ability to participate in meaningful forms of work. Achievement gaps are interesting, in part, because they are the inverse of the (negative) responsibility gaps already widely discussed in the literature on AI ethics. Having described and explained the problem of achievement gaps, the article concludes by identifying four possible policy responses to the problem.

Keywords: Automation, achievement, meaningful work, artificial intelligence, responsibility gap, autonomy, mastery, community.

1. Introduction

On the 27th of November 2019, Lee Sedol announced his retirement from the professional world of Go-playing.¹ Sedol was an eighteen-time world champion of the fiendishly difficult Chinese strategy game. He was a long-time celebrity in his home country of South Korea but attained international notoriety in March 2016 when he was defeated, 4-1, in a five-game challenge match against *Alpha Go*, an artificially intelligent Go-playing computer program, created by the software company DeepMind.

Sedol did not announce his retirement with a sense of satisfaction about what he achieved in his career. He did so out of a sense of despair. He told reporters that he felt there was no point in continuing with the game because he could no longer compete with the machines. He had come to the realization that

¹ For Lee Sedol's retirement and associated statements, see: <https://www.bbc.com/news/technology-50573071>

he was no longer "at the top even if [he became] the number one".² Sedol's sense of despair at the rise of artificial intelligence is shared by others. Commenting on Sedol's loss to *Alpha Go*, Carissa Veliz notes that there is something deeply sad about the whole affair. Clearly, the creation of Alpha Go marked a significant technological breakthrough, but it also signalled the loss of something important to humankind. As she put it:³

"What is most surprising about the match is that the outcome did not feel like a win for humanity. It did not feel similar to when we conquer a disease, or when the first human being landed on the moon. It felt like we might be losing more than what we might be gaining."

In the remainder of this article we examine what it is we might be losing to machines like Alpha Go. We focus, in particular, on what we might be losing in the world of work. As AI and other automating technologies are increasingly used to augment and replace human task performance in the workplace, there is, we argue, a significant risk to the value of human achievement. This, in turn, makes it difficult for us to ensure that people have access to meaningful work. And while this argument focuses specifically on the impact of automation in the workplace it may have broader implications for the impact of automation on human life more generally.

We present our analysis in four stages. First, we clarify in a bit more detail the phenomenon of workplace automation and the idea of meaningful work. Second, we look at the nature of human achievement and identify four conditions that need to be satisfied in order for us to say that a human or group of humans has achieved some goal or end. Third, we argue that workplace automation undermines our capacity to individually and collectively satisfy these conditions for achievement, thus resulting in an 'achievement gap' in the automated workplace. Fourth, we consider the policy implications of this argument. In particular, we outline ways of retaining a space for human achievement in the automated workplace, or otherwise guaranteeing meaningfulness in the workplace. In doing so, we draw explicit analogies between our concern about the emergence of achievement gaps in the workplace and closely related policy concerns about responsibility gaps arising from the widespread use of automating technologies (e.g. Mattias 2004; Nyholm 2018 and 2020; and Tigar 2020a and 2020b). Indeed, we suggest that the achievement gaps we discuss in this paper are a type of positive responsibility gap, and are thus the inverse of the negative responsibility gaps that have attracted a lot of attention in debates about roboethics and AI ethics.

² *Ibid*

³ These remarks come from a blogpost Veliz wrote after watching a documentary about the Alpha Go story, available at:

https://blog.practicaethics.ox.ac.uk/2019/10/a-sad-victory/?fbclid=IwAR1CIVsJI3qm4fW-TmNOjN67x7UR_fNjkr1akVqS7DZC804pzqUhhBMBpgQ

2. Meaningful Work and Workplace Automation

This paper is concerned with the impact of automation on our capacity for achievement in the workplace. To better understand both the scope and limitations of our analysis it is important to clarify what we understand by the phenomena of work and automation, and then to consider the relationship between achievement and meaningful work.

Let's start with the idea of work. Work is a tricky phenomenon to define (Danaher 2019, ch 2): some people adopt expansive definitions of work, defining it in such a way that it includes virtually all activities that humans might engage in; others adopt more limited definitions, tying work to specific, economic activities. For present purposes, we use one of these more limited definitions. We focus on what has been called work 'in the economic sense' (Gorz 1989; Frayne 2015). This form of work consists of skills (physical, cognitive, emotional etc) that are performed by individuals in return for some kind of economic reward. The skills performed by the individuals can be referred to as work-related 'tasks'. Usually, these tasks are performed with a view to producing a commodifiable output, either a good or service, that can be bought and sold on a market. Workplaces then consist of individuals, either singly or collectively, performing work-related tasks with a view to producing such commodifiable outputs.

Work, so-defined, often has a nested task structure. That is to say, one set of work-related tasks can produce an output that then contributes to another set of tasks that contributes to the overall commodifiable output. For example, in the manufacture of a complex technological product like a computer, one group of workers may be focused on producing designs or plans for the computer, another group may be focused on manufacturing the microchips, another on the casing and packaging, and still another may be dedicated to marketing and selling the product. Each group of workers produces an output that contributes to the overall commodifiable output. The nested task structure is due to specialization within the workplace, a phenomenon highlighted long ago by Adam Smith (Smith 1776). The nested task structure of work is important when it comes to discussions of achievement in the workplace because the output of a given task (or set of tasks) is one of the key variables when it comes to assessing the value of the achievement associated with the work. A worker may, for example, be contributing to some valuable commodifiable output through their individual work tasks, but the outputs associated with their work tasks may be relatively trivial or easy to achieve. If this is the case, their particular work task may not be a significant source of achievement. To take an extreme example, a legal secretary whose only job is to staple together sheafs of paper is doing something that is relatively trivial and easy to achieve. It is not clear how their job could be a significant source of achievement (though we return to examples like this, again, in section 5).

Work, so defined, is not the same thing as a job though they are related concepts. A job is a defined role within a workplace associated with the performance of a more or less specified collection of work-related tasks. Put another way, work is a general condition under which tasks get performed

(namely, a condition of economic reward) whereas a job is a collection of such tasks associated with a specific workplace identity. These tasks must be connected to the commodifiable output, but apart from that they can be more or less arbitrarily defined. For example, one person might have the job of being a personnel manager within a workplace, but the precise set of work-related tasks relevant to this job might shift and change over time. At one point in time the major focus might be on providing adequate skills-training to the workforce, at another point in time the major focus might be on improving workplace morale and so on. This is important because jobs can often be redefined or altered over time to include different sets of work-related tasks. This makes a difference when it comes to assessing the impact of automation on work. As we point out below, automation may result in a redrawing of the task boundaries of particular jobs without necessarily resulting in a net loss of jobs.

Finally, in relation to the definition of work, it is worth distinguishing between meaningful and non-meaningful forms of work. 'Meaningful work' is work that consists of a set of tasks that provides workers with meaning, where this is cashed out in terms of the general properties commonly associated with meaning in life. This could prompt a lengthy divagation into the nature of meaning in life but we will take a shortcut around that topic in this paper. In our approach to meaning in work we are influenced by Susan Wolf's account of what it is to have a meaningful life. Meaning in life, according to Wolf (2010), has both a subjective component and an objective component. As Wolf sees things, a person has a meaningful life if she is engaged in activities and projects that she is passionate about, and that can also be recognised as valuable from a wider, not purely subjective perspective. We think of meaningful work in a similar way: it should create a subjective sense of meaning, on the one hand, but it should also be work that has characteristics that can be more generally recognised as the sorts of characteristics objectively associated with meaningful work, on the other hand (for an extended discussion of such characteristics, see, e.g., Smids, Nyholm, & Berkers 2020, pp. 506-510).

What does this mean in practice? At a minimum, two conditions need to be satisfied in order for work to be meaningful: (a) the overarching output produced by the workplace must be valuable and must be perceived by the workers to be valuable and (b) the individual worker's job, within the overarching structure of the workplace, must consist of sub-tasks that are themselves valuable (and perceived to be valuable). So, for example, someone could work as a financial trader within a workplace that specializes in facilitating illegal arms trades that ultimately harm many people, or, alternatively, they could work as a trader that facilitates investments in healthcare in developing countries that benefit many people. In both cases they might perform very similar tasks as part of their jobs and find these tasks to be meaningful and valuable, but when facilitating illegal arms trades they are not engaged in a meaningful form of work because the overarching output of their workplace lacks value (indeed, it is actively disvaluable). All that said, as we shall point out later in this paper, meaningful work can also include other properties not specifically linked to the output of the workplace (Ghaeus and Herzog 2016; Smids, Nyholm and Berkers 2020; Veltman 2016). For example, feeling as though

you belong to a community of fellow workers, or that you are gaining mastery over some skillset, are often tagged as features of meaningful work, even though these things do not strictly require a valuable task output – they require, at most, a value neutral output.

So much for work. What about workplace automation? Very generally, automation is the practice of using machines, rather than humans, to perform work-related tasks. Automation has been a feature of the workplace since the dawn of the industrial revolution (Frey 2019). Some people think that the current wave of AI-based automation could eventually lead to the end of the majority human workforces (Ford 2015; Brynjolfsson and McAfee 2014; Manyika et al 2017; Danaher 2019; Susskind 2020); others are more skeptical (Autor 2015). This is not the place to adjudicate those disputes. What is important for present purposes is that workplace automation can take different forms, two of which are worth singling out for the purposes of this article:

Total Replacement: This happens when a machine replaces all the work-related tasks associated with a given job. As a result, a human worker will typically lose their job and a machine will perform all the work-related tasks that were once associated with that job.

Collaborative Displacement: This happens when a machine replaces some of the work-related tasks associated with a given work output but a sufficient number of human-performed tasks are retained such that humans still have jobs but have to collaborate with the machines to produce the output. This often leads to a redrawing of the boundaries of a given job.

Collaborative displacement, in its turn, can take a number of different forms. It would be impossible to enumerate all the possible forms it can take but three would seem to be particularly important when it comes to assessing the impact of workplace automation on achievement.

The first is what might be called “supervisory” or “directive” collaboration. This arises whenever humans formulate the abstract plans or work programs that the machines implement. For example, a group of automobile engineers might design a car, providing detailed blueprints for its specifications and information on how it ought to be put together. These blueprints could then be fed into a machine (or group of machines) that actually build the car. The humans here are responsible for the creative and intellectual part of the work. The machine does the physical gruntwork.

The second is what might be called “maintenance” collaboration. This arises whenever humans have to maintain the machines that produce the work output. To continue with the automobile example, once the machines have started the manufacturing process, it is quite likely that they will breakdown or encounter operational difficulties. Skilled machinists and maintenance workers may then be required to step in to repair or fix the machines, or reprogram/repurpose them to fit the changing demands of the workplace.

The third is what might be called “order-following” collaboration. This is essentially the inverse of “supervisory” or “directive” collaboration. In this case, the machine comes up with the abstract plans or work programs that human workers then implement. So, suppose that instead of designing the car, the humans use a machine learning algorithm to figure out the best way to make a fuel efficient, aerodynamic car. The algorithm uses a database of pre-existing designs to do this. After the algorithm settles on a preferred design, the humans go off and build the car. Here, the humans do the physical gruntwork while the machine is responsible for the creative and intellectual work.

These examples are obviously abstract and, perhaps, a little too neat. In the real world, multiple forms of collaborative displacement may occur at the same time and some may be more partial or incomplete than those sketched above. Nevertheless, the different forms of collaborative displacement make a difference to the kinds of achievement that are available to human workers. Humans might still have jobs, even with rampant automation, but those jobs might not allow for meaningful work because the kinds of achievement they enable are relatively minimal. The obvious intuition here is that a human that just follows the orders given to it by a machine has access to a less valuable kind of workplace achievement than one that supervises or directs a machine. But is that intuition correct? We will examine this in more detail below, after we have said something more about the nature of achievement.

3. What is an achievement?

Achievements are outcomes that are brought about by coordinated human activity in such a way that those outcomes can be linked to the efforts of individual human agents. The concept of achievement features heavily in many philosophical theories of meaning in life and hence can also feature in theories of meaningful work. A meaningful life is commonly conceptualized as one in which an individual achieves valuable things with their life. Albert Einstein, for example, can be said to have lived a meaningful life because he achieved certain scientific insights and outputs (the special and general theories of relativity and certain key elements of modern quantum theory) through his cognitive efforts. That said, his life was not an unqualified success since he failed to achieve a grand unified theory of everything, something that both he and the wider physics community perceived to be the ultimate goal of their scientific endeavours (Folsing 1997).

Achievement is, then, one of the key elements of a meaningful life. It may not be the sine qua non of human flourishing but it is at least true to say that a life without achievement is impoverished compared to a life with achievement. If something threatened or undermined our capacity to have achievements, we should be concerned as this could make our lives relatively worse than they might otherwise have been. This is particularly true when it comes to our desire to secure meaningful forms of work. For better or worse, work occupies the centre stage in many people’s lives. They are obliged to do it out of economic and practical necessity; and they often associate their identity and self-worth with

the work they perform, and many people spend more time every week with workplace colleagues than with their families and loved ones (Smids, Nyholm and Berkers 2020). If their capacity to achieve things through their work is compromised, this would be worthy of careful consideration.

What conditions are relevant to assessing whether someone's efforts count as an achievement? Here we propose a modified account based on Bradford's analytical account of achievement (Bradford 2013 and 2016) and Maslen, Savulescu and Hunt's work on praiseworthiness (2020). It seems appropriate to marry these two accounts together since achievements are, in essence, a positive manifestation of responsibility. In both philosophical and legal debates about responsibility, we are often concerned with what can be called negative responsibility: who is to blame (legally/morally) when something bad happens? It is important, however, to remember that responsibility has a positive, flipside. We can also ask: who deserves praise or reward if something good happens? Since we usually talk about an individual's achievements in positive terms – e.g. we celebrate Einstein's scientific achievements – it makes sense to think that assessing whether or not someone has achieved something is partly determined by whether or not they deserve praise for a positive outcome.

Some people might dispute this. Steven Luper (2014), for example, has defended an "achievementist" theory of meaning in life that is strictly neutral on whether achievements have objective value. He argues that someone can be rightly said to achieve negative outcomes just as easily as they can be said to achieve positive outcomes. Achievement, according to Luper, is largely a function of whether or not someone is successful in attaining their goals. To take an extreme example, a serial killer who set themselves the goal of killing as many people as possible, and who succeeds in doing this, can be said to have achieved something just as easily as Einstein can be said to have achieved something through his scientific successes.

While there might be some merit to such an account of achievement, we suggest that it is a counterintuitive idea. If achievements are a key part of meaning in life, it would be odd to suggest that they can play this part if they can be negative in nature. Accordingly, it is common in theories of meaning to argue that a life lacks meaning if it produces negative outcomes in the world (Landau 2011; Smuts 2013; Metz 2013). Because we find Luper's view to be counterintuitive, in the remainder of this article, we focus on achievements that are associated with either positive or non-negative outcomes (i.e. neither good nor bad). We include the latter possibility on the grounds that achieving a neutral non-negative outcome can still be praiseworthy because (a) it doesn't reduce the level of value in the world and (b) the fact that it was achieved may speak to certain virtues or excellences in human agency. For example, it seems plausible to suggest that Lee Sedol's successes as a professional Go-player were largely value neutral in nature: his winning the world championship 18 times was, from an objective standpoint, neither good nor bad. Nevertheless, his successes did speak to some personal virtues or excellences that are worthy of praise.

With these caveats in mind, we submit that whenever we are assessing the value of someone's achievements, the following four variables need to be kept in mind:

- (a) *The value of the output produced*: although producing value neutral outcomes can count as an achievement, it is still the case that the more objectively valuable the output then the greater the achievement, all else being equal (i.e. bearing in mind the other variables).
- (b) *The nature of causal contribution of the agent*: the agent must play a significant role in producing the output, specifically the output must be a sufficiently *non-lucky* result of the agent's actions.
- (c) *The cost of the agent's commitment to producing the outcome*: typically, the more effort, time, attention, stress etc. the greater the achievement.⁴
- (d) *The voluntariness of the agent's actions*: a non-voluntary commitment to producing an outcome might count as a less of an achievement, all else being equal.

The first of these variables has already been discussed. The other three merit some further elaboration. Concerning the causal contribution of the agent to the output, it seems like commonsense to suggest that this must be significant and that the agent has to play a sufficiently 'non-lucky' role in producing the outcome (Bradford 2013). To use an obvious example, compare Bill Gates to a lottery winner. The lottery winner purchases a ticket and this ticket happens to have the winning numbers. This results in the lottery winner becoming fantastically wealthy. Did the lottery winner achieve their wealth? Hardly. Their purchasing of the ticket was a relevant causal factor in producing this outcome (*but for* the purchase of the ticket they would not be wealthy) but it was pure luck whether this ticket would actually be the winning one. The lottery winner did not play a sufficiently non-lucky role in producing the outcome. Bill Gates, on the other hand, was undoubtedly the recipient of much good fortune and luck in life. He came from a relatively wealthy background and happened to come of age at the exact right time to capitalize on the personal computer revolution. Nevertheless, he did exert considerable, deliberate cognitive effort in programming key software innovations and managing a successful business (Isaacson 2014). His

⁴ As will become apparent, we do not think that an agent's exercising a great amount of effort is a necessary condition for something to count as an achievement on the part of the agent. Assuming, for example, that Mozart was able to compose an opera with a smaller amount of effort than some lesser composer, this may not necessarily take away anything from how great of an achievement the opera might be on Mozart's part. In general, however, we typically associate achievements with exertions of effort or other expenditures of time and attention (etc). Something's having required a certain amount of effort is one of the sorts of things that are often associated with its being worthy of being considered an achievement.

resultant wealth may be excessive relative to what he deserves, but he did play a sufficiently non-lucky role in acquiring that wealth.⁵

The cost of the agent's commitment to the outcome also seems obviously relevant to the assessment of their achievements. As Maslen et al (2019) point out in their discussion of praise, it is common for people to think about the cost of commitment in terms of effort. But what is effort? Maslen et al define it as a psychological capacity to overcome aversive experiences. It takes a lot of effort, for example, to run a marathon. It is a physically and psychologically demanding thing to do. Many times, as you are running, your body and mind will want to give up. You have to exert effort to overcome that desire to quit. Understood in these terms, effort is clearly relevant to assessing achievements. The more effort expended on producing an outcome, the greater the achievement seems to be. That said, as Maslen et al argue, it is also quite a limited way of looking at things. Effort isn't the only way for an agent to commit to producing an outcome and sometimes effort may not be present in significant achievements. Compare, for example, an amateur first-time marathon runner to a professional. The amateur may have to exert far more effort (overcome more psychological aversion) than the professional to finish the race. This does not, however, make the amateur's marathon a greater achievement. The professional will have shown their dedication to the race in other ways. They will have spent many hours training and honing their skills; they will have spent time away from their families and friends in order to improve their abilities. These are all forms of costly commitment and they all seem relevant to assessing the achievement. That said, we have to treat costly commitment with some degree of care. It is relevant to the assessment of an achievement but not necessary for it. As Maslen et al point out, for costly commitment to be relevant to the assessment it has to be non-arbitrary. In other words, you cannot just make something arbitrarily more difficult in order to increase the amount of praise you are owed. In fact, sometimes, finding an easier, less costly way to produce the same outcome, can be an achievement in and of itself. We return to this point later on when we assess the impact of automation on achievement.

The voluntariness of the agent's commitment to the outcome also seems like something that is obviously relevant to the assessment of an achievement because it signals a positive desire to achieve an outcome. Let's return for a moment to Bill Gates. As is well-known, Gates has pledged to give most of his wealth away to charitable/good causes through the foundation he set-up with his wife. As far as we are aware, this pledge was entirely voluntary. No one forced him to do it. But suppose this were not true. Suppose it turned out that someone had held a gun to his head and demanded that he give away all his money. Would this make a difference to our assessment of the pledge? It might. The fact that he only gave away the money under significant pressure would seem to suggest that he doesn't really deserve as much credit for it. That said, as with costly commitment, we have to be careful in how we factor voluntariness

⁵ And, just to be clear, we are not claiming that Bill Gates' amassing of that wealth is objectively valuable. We use him simply as an example of what it might mean to be a sufficiently non-lucky producer of an outcome.

into our assessment of achievements. Just because an action is non-voluntary doesn't mean it doesn't count as an achievement. Someone might force an amateur to run a marathon at gunpoint. If the amateur runs the marathon (and completes it) this would still count as a significant achievement, given the degree of psychological effort it must have required to do so. The point, then, is that voluntariness is not the be all and end all of achievement. It is one of the factors that goes into the assessment. You must reach a holistic assessment of all four factors to determine whether there has been a significant achievement or not.

4. How Automation Creates Achievement Gaps

With this account in place, we are in a better position to assess the impact of workplace automation on achievement. Instead of asking the general question about how automation impacts on achievement we can ask the more specific question: how does automation impact on the four variables relevant to the assessment of achievements? Answering this will then enable us to reach a holistic assessment. To answer this question adequately, we need to link the discussion back to the different kinds of automation discussed in section 2.

We can start with the trivial case: that of total replacement of human workers by machines. Clearly, in such cases, there is an end to certain kinds human achievement within the workplace. If a skilled manufacturing worker or service worker is replaced, entirely, by a machine, they no longer have access to any form of workplace achievement. Whatever achievements they used to associate with their job are now closed to them: they lose the ability to make a non-lucky causal contribution to a valuable output. They have to find other outlets for their talents. This is an uncomplicated and perhaps uninteresting case (though it leads to a complicating factor that we discuss in more detail below).

The more interesting cases are the ones in which workers are not totally replaced by machines but are, instead, redeployed within the workplace so that they have to collaborate with machines in different ways. In all cases of collaboration, human workers perform tasks that (a) produce outcomes in their own right and (b) have some causal relevance, even if indirect, to the outcomes produced by the machines (and therefore of the workplace as a whole). There is, consequently, a valid question to be asked about whether they still have access to achievement through their work. What we will now argue is that even in cases of collaborative displacement, the possibility of achieving things through your work can be significantly compromised through the use of automating technologies. This is because even in cases of collaborative displacement automation can negatively impact on three of the four variables that are relevant to the assessment of achievements in the workplace.

Consider, first, the way in which automation can reduce the value of the outputs associated with the tasks performed by certain groups of human workers. The clearest example of this comes with maintenance forms of collaboration. Where once upon a time humans could be responsible for producing valuable commodifiable outputs, either through cognitive design of

those outputs or physical manufacture of their parts or, in the case of services, physical and cognitive performance of valuable services, when they shift to the maintenance role they necessarily take a peripheral role in the production of the valuable outputs. They keep the wheels of industry turning, but they do not play a direct role in the valuable activities. Consider, by way of analogy, the relationship between a touring musician and their road crew. The road crew plays an important part in setting up and testing the equipment that ensures the musician can play every night, but they are playing a support role to the activity that produces the real value – the musical performance. You would be hard pressed to argue that their tasks are *more valuable* than those of the musician. In the case of maintenance collaboration, human workers are relegated to an equivalent backstage role. They are the support crew not the performers. What they do may still be causally important, but it is not more valuable than the frontstage role they could have occupied.

Similarly, in cases of order-following collaboration, it is reasonable to presume that there will be a loss of value in work-related outputs, relative to other, available forms of work. In the case of order-following collaboration, humans take a backseat to the creative, design or supervisory work. They become order-following, physical workers. The tasks they perform may well be causally important to the production of valuable outputs, but what they do is surely relatively less valuable. Why? Because in becoming order-following physical workers they lose the opportunity to exercise independence and creativity in their jobs. Indeed, there is something deeply ironic about order-following collaboration. As Brett Frischmann and Evan Selinger (2018) have argued, one of the more noticeable effects that automation and digital technology has on human life is its tendency to reduce humans to ‘simple machines’ that perceive a stimulus and then respond in a predictable or limited way. Think, for example, of the Amazon warehouse worker who is given a set of customer orders and a route through the warehouse to pick up those orders that has been planned by an algorithm. They see the stimulus given to them by the machine and respond, not with creativity or foresight, but by just following the route that is recommended. They dare not second-guess the algorithm or exercise any creativity in case they are less efficient at their jobs (and are reprimanded for failing to follow the orders). Similar things are happening in other forms of work where algorithms play a key role in planning and coordinating the physical activity of workers, e.g. in ‘platform’ work such as that provided by companies like Uber and Deliveroo.

The one major exception to this argument is the case of supervisory collaboration. Here, the humans would appear to retain creative control and mastery over the machines. They still use their cognitive abilities to design the valuable output and supervise its creation. While they may have outsourced the physical gruntwork to a machine, they still play a significant, non-lucky role in producing that output. But even in the case of supervisory collaboration there can be a loss. For one thing, it is worth noting that creative and supervisory forms of work are largely the preserve of an elite few workers and, indeed, one of the tragedies of automation may well be that it empowers these elite few workers to the disadvantage of the majority (Autor 2015; Loi 2015; Frey 2019).

Related to this, even in cases of supervisory collaboration, there can be a loss in the value of work-related outputs, depending on the form of automating technology in use. To return to the opening example of AlphaGo and Lee Sedol, one of the key features of the machine learning system used by the designers of AlphaGo is that it was largely independent of, and opaque to, its designers. In other words, AlphaGo learned how to play the game by itself, by extrapolating strategies and tactics from previous games, that were surprising to both its creators and professional Go-players. Thus, while the programmers did create the machine and supervised its use, they were not, in any meaningful sense, responsible for its specific successful moves within the game. If this form of automating technology becomes more widely used it can undermine the value of even supervisory forms of collaboration.

This links to a second important argument about automation and collaborative displacement, namely: that automation, almost by necessity, reduces the cost of the human commitment to producing a workplace output. In some ways, this is the primary motivation for using automating technologies. They are supposed to reduce the need for certain forms of human effort and activity, and they are usually cheaper and more efficient than human workers in doing so (this is often the primary economic rationale for introducing them). But if costly commitment is one of the key variables for assessing the value of an achievement it would seem to follow that automation reduces the value of workplace related achievements for workers. The designers of AlphaGo would again appear to be a case in point. They did not have to do the hard work of figuring out the optimal strategies to play in the game, nor did they have to code these strategies into the computer. Something similar is true for the worker who follows orders given to them by a machine. They do not have to do the hard work of planning routes or deciding on what needs to be done. They just implement the orders. Think about the Uber driver. Unlike a traditional taxi driver, they do not need to drive around a city looking for potential customers. The software directs the customers to them and them to the customers. With the addition of mapping software, it even plans the driving route for them, and facilitates payment. This significantly reduces the cost of the commitment to producing the work output.

There is, of course, an obvious counterargument to this. As noted earlier, costly commitment comes in a variety of forms and it is important to remember that even if one type of cost is reduced, other forms of cost could go up and compensate for this. In the earlier example of the marathon runner, we suggested that although the effort expended on each individual race might be lower for a professional runner than for an amateur runner, the professional will demonstrate their costly commitment in other ways (tougher, longer training regimes and so forth). Could something similar arise in the case of workplace automation? Could the automation of one type of work task increase costly commitment elsewhere in the workplace? This sounds plausible. It is not unusual to hear people complain that so-called 'labour saving' technologies have actually increased the demands on their work in other ways. So, for example, the Uber driver might argue that although the technology saves them from having to find customers and plan driving routes, they spend an increased amount of time

driving in order to earn a decent wage and have to invest more energy and effort in improving their customer service to boost the ratings they receive from passengers (Rosenblat 2019).

There are several things to be said in response to this. First, as we noted earlier, costly commitment is not always and everywhere associated with an increase in the value of an achievement. Sometimes costs can be arbitrary or counterproductive. It could be that the additional costs imposed by workplace automation are of this form. For example, the Uber driver may have to spend more time driving because one of the effects of platform markets is reduced wages and job security (Weil 2014; Manyika et al 2016; Rosenblat 2019; Prassl 2019). So the increased commitment is, in fact, an indication that the form of work has become less valuable in other ways. Second, even if costs are not arbitrary or counterproductive, we still have to ask where the additional costs are directed. Our previous argument held that one of the effects of workplace automation is a reduced value of the outputs associated with human work tasks. Increasing the cost of commitment to less valuable outputs would still amount to a net loss in the value of achievements associated with work. For example, the Amazon worker who expends significant physical effort running around a warehouse in order to fulfill customer orders would still seem to be losing out, relative to what might have been the case, because the activities they perform are less valuable in the overall scheme of things.

This brings us to another argument about the impact of automation on achievement. A non-lucky causal contribution to an output is, as noted earlier, a relevant variable when it comes to assessing achievements. But again, almost by necessity, one of the things that automating technologies tend to do is to sever the causal connection between human activity and outputs. This is because automating technologies are introduced primarily to shift labour (physical or cognitive) from humans to machines. The classic case is that of the manufacturing robot that takes over the physical tasks of producing an industrial output such as an automobile. This manufacturing robot severs the connection between human workers and that output.⁶ The result is that the human workers can no longer claim a direct, non-lucky, causal role in producing the output that was previously earmarked as their achievement. This severing of the causal connection could be compensated for if the humans were given other, more valuable, things to do but, as per the previous arguments, this is may not be true for the majority of workers.

What about the voluntariness condition on achievement? Could the rise of automating technologies in the workplace somehow compromise the voluntariness of work? It is hard to see how this could be the case, unless the machines become, in effect, slave-masters of humans.⁷ That said, it should be

⁶ This brings to mind Marxist – and related – ideas of alienation – something we think would be worth exploring further, but that we will leave for some other occasion. For an overview of such ideas, see Leopold 2018.

⁷ For relevant discussion of how some contemporary technologies might be thought to make people more like machines who follow instructions in a mind-

noted that voluntariness of work is already compromised for many workers (Danaher 2019; Anderson 2017). Unless you are fortunate enough to have personal wealth that obviates the economic necessity of work, or have plenty of viable career options that mean you are not forced to choose among a narrow range of potential jobs/work tasks, your ability to freely choose work is already quite limited. Automation may exacerbate this, to some extent, by further narrowing the pool of viable options, and so automation may put pressure on the voluntariness of work, but we don't rest too much weight on this here in making the case for achievement gaps since, as already noted, voluntariness does not necessarily undermine achievements.

In sum, the introduction of automating technologies into the workplace has the potential to open up numerous achievement gaps. This is because automating technologies tend to reduce the value of the outputs associated with human work tasks, reduce the cost of the human commitment to producing valuable commodifiable outputs (or redirect costly commitment in arbitrary and counterproductive ways) and, ultimately, sever the causal connection between human workers and valuable outputs. This is a bad thing since achievement is a key component of meaningful work. The challenge from a policy perspective is to figure out what can be done to mitigate against this risk. We turn to that question next.

5. What can we do about achievement gaps in the workplace?

In this section, we sketch four possible responses to the achievement gap problem in the workplace. We emphasise that these are just *sketches*: things that might be done to alleviate some of the problems discussed in the previous section. More work needs to be done to develop a comprehensive policy response. Nevertheless, we can give some guidance on the type of thing that should be done.

The first possibility is to simply accept the threat to achievement and respond by emphasising other aspects of meaningful work. Earlier on, when discussing the phenomenon of meaningful work, we noted that achievement, while important, is not the only thing that makes work meaningful. Leading accounts of meaningful work suggest that it is a complex and multidimensional phenomenon. Meaningful work includes achievement of valuable outputs, to be sure, but it also includes a sense of community, gaining self-respect and recognition of one's peers, mastering skills and having a high degree of autonomy in the performance of one's work tasks (Smids, Nyholm and Berkers 2020; Gheaus and Herzog 2016). While automation may reduce the significance of the achievements associated with one's work-related tasks, we could

less way, and who might potentially therefore be viewed as having these technologies as their masters, see Frischmann & Selinger 2018. For the idea of the master-slave dialectic in the context of human-technology interaction with a Hegelian twist, see also Coeckelbergh 2015.

compensate for this by trying to promote the other aspects of meaningful work. For example, the employee that is displaced by a machine could be given greater autonomy in their remaining work tasks, workplace training and performance management could focus on enabling them to master the skills associated with these new tasks, greater emphasis could be placed on workplace morale and camaraderie and so on. That said, this strategy does face two obvious problems. First, emphasizing these other aspects of meaningful work may ring hollow if an individual worker's contribution to the workplace output is reduced. Sure, it's nice to have more autonomy and good workplace moral, but what's the point if all one is doing is stapling sheets of paper together (to continue an example from earlier in the paper)? Second, increased workplace automation may put pressure on other aspects of meaningful work too. For example, one may have fewer colleagues – and hence less of a sense of community – in a highly automated workplace. Likewise, ensuring that humans work smoothly with machines might require more worker surveillance and control, due to the risk that humans will misuse machines or machines pose a risk to humans. This may, in turn, impact negatively on autonomy.

The second possibility is that, even if machines play the main part in producing workplace outputs, we could find some residual ways for humans to play a valuable role in producing the final outputs. To put it another way, we could try to find a way for the 'human touch' to be retained in the automated workplace. For example, imagine a furniture factory that relies heavily on machines to manufacture tables and chairs. The machines produce safer and cheaper furniture than human workers ever could. Suppose though, at the very end of the manufacturing line, human workers add distinctive carvings or markings to the furniture. These workers are given a high degree of freedom and autonomy in doing this and hence are able to exercise their creativity when it comes to marking up the furniture. This could add something distinctive to the final product and allow human workers to play a meaningful role in production. Of course, work is not charity. We cannot simply dream up tasks just so that all products or service have some semblance of the human touch. But adding such distinctive markings to furniture may not be charity. Consumers may value a product that has unique features and includes the human touch.⁸ It may counterbalance some of the drab uniformity that is often associated with machine-manufactured goods. Indeed, setting up the workplace in this way may enable us to get the best of both worlds: the efficiency of machines and the creative imperfections of humans.

The third possibility is that we place greater emphasis on the role of teams in producing workplace outputs. This isn't really a way of addressing the problems outlined in the previous section but, rather, a way of reframing them so that they seem less problematic. One criticism that could be made of our arguments in the previous section is that they adopt an overly individualistic

⁸ An anonymous reviewer points out to us that there is, in fact, a furniture company – Sticks, Inc – whose unique selling point is that every piece they sell is finished with hand-drawn artwork. See <https://sticks.com> for more details. We thank the reviewer for this reference.

understanding of achievement and meaningful work: in order for *my* work to be meaningful *I* must play a significant role in producing the valuable output. This was perhaps most clearly on display in the dismissive comments we made about support crew and stage performers. But maybe this is the wrong way of looking at it? Many of the great achievements of human history have been collective efforts not solo runs. No person is an island unto themselves. We all depend on and rely on others to get things done. If any link in the collective chain gets damaged, the final product suffers. Maybe we just need to reflect on that reality more and accept that collaborative displacement by machines doesn't undermine achievement as much as we supposed. Perhaps we can still feel like we are a valuable part of the team producing the output, even if the tasks we perform are different? Perhaps. But, again, this might ring hollow to those who find themselves redeployed to more marginal and peripheral roles within the team. The tasks they perform will need to feel somehow essential to the collective effort (a necessary part of a sufficient set of causes of the output).

The final possibility is that we try to find other avenues for achievement outside the workplace. This may be the ultimate solution to the problem. The argument developed in this paper has worked from the assumption that work plays a central role in people's lives and that it is one of the main forums in which they can achieve valuable outputs. This may be true given the current economic realities but maybe it doesn't have to be true in a world of widespread workplace automation (Danaher 2019)? With fewer demands on humans in the workplace it may be possible for us to dedicate more time to other methods of achieving valuable outputs, e.g. hobbies, sports, artistic endeavours, family projects and so on. This links, importantly, to the voluntariness condition for praise. One of the problems noted above is that work is often not a voluntary choice for many people: it is an economic and practical necessity. But if automation could free us from that necessity, we might be able to choose our own paths to achievement. That said, one of the features of the argument we made in the previous section is that it is generalizable. We focused on achievement gaps in the workplace but, in principle, the use of automating technologies in other domains may open up additional achievement gaps. This may reduce the viability of this final strategy (cf Danaher 2019, ch 4 and ch 7).

6. Conclusion

Achievement is an important part of the well-lived life. It is the positive side of responsibility. Where we blame ourselves and others for doing bad things, we also praise ourselves for achieving positive (or value neutral) things. Achievement is particularly important when it comes to meaningful work. One of the problems with widespread automation is that it threatens to undermine at least three of the four main conditions for achievement in the workplace: it can reduce the value of work tasks; reduce the cost of committing to those work tasks; and sever the causal connection between human effort and workplace outcome. This opens up 'achievement gaps' in the workplace. There are, however, some potential ways to manage the threat of achievement gaps: we can focus on other aspects of meaningful work; we can find some ways to retain the human touch in the production of workplace outputs; we can emphasise the

importance of teamwork in producing valuable outputs; and we can find outlets for achievement outside of the workplace.

Acknowledgements: Nyholm's work on this paper is part of the research program "Ethics of Socially Disruptive Technologies", which is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.004.031).

Conflicts of interest/Competing interests: None

Availability of data and material N/a

Code availability: N/a

Authors' contributions: 50/50

Bibliography

Anderson, Elizabeth (2017). *Private Government*. Princeton, NJ: Princeton University Press.

Autor, D. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives* 29(3): 3-30.

Avent, R. (2016). *The Wealth of Humans: Work, Power and Status in the 21st Century*. New York: St Martin's Press

Black, B. (1986). *The abolition of work and other essays*. Port Townshend, Washington: Loompanics Unlimited

Bradford, G. (2013). The Value of Achievements. *Pacific Philosophical Quarterly* 94(2): 202-224

Bradford, G. (2016). *Achievement*. Oxford: Oxford University Press.

Breemersch, K., J. Damijan and J. Konings (2017), "Labour Market Polarization in Advanced Countries: Impact of Global Value Chains, Technology, Import Competition from China and Labour Market Institutions", OECD Social, Employment and Migration Working Papers, No. 197, OECD Publishing, Paris, <https://doi.org/10.1787/06804863-en>.

Brynjolfsson, E. and McAfee, A. (2014). *The Second Machine Age*. New York: WW Norton and Co.

Coeckelbergh, M. (2015). "The Tragedy of the Master: Automation, Vulnerability, and Distance", *Ethics and Information Technology* 17(3): 219-229

Danaher, John (2019). *Automation and Utopia: Human Flourishing in a World Without Work*. Cambridge, MA: Harvard University Press.

Folsing, Albrecht (1997). *Albert Einstein: A Biography*. New York: Viking

Ford, M. (2015). *The Rise of the Robots: Technology and the threat of Mass Unemployment*. New York: Basic Books

Frayne, D. (2015) *The Refusal of Work*. London: ZED Books.

Frey, CB. (2019). *The Technology Trap*. Princeton, NJ: Princeton University Press.

Frischmann, B. and Selinger, E. (2018). *Re-engineering Humanity*. Cambridge, UK: Cambridge University Press.

Gheaus, A., and L. Herzog. 2016. The goods of work (other than money!). *Journal of Social Philosophy* 47(1): 70–89. doi: 10.1111/josp.12140

Gorz, André (1989). *Critique of Economic Reason*. London: Verso

Isaacson, Walter (2014). *The Innovator*. London: Simon and Schuster.

Landau, Iddo (2011). Immortality and the Meaning of Life. *Journal of Value Inquiry* 45: 309-317

Leopold, D. (2018). "Alienation", *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/fall2018/entries/alienation/>>.

Loi, M. (2015). Technological Unemployment and Human Disenhancement. *Ethics and Information Technology* 17:201-210.

Luper, S. (2014). Life's Meaning. In Luper, S (ed) *The Cambridge Companion to Life and Death*. Cambridge, UK: Cambridge University Press.

Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P. and Dewhurst, M. (2017) *A Future that Works: Automation, Employment and Productivity*. McKinsey Global Institute Report, McKinsey and Co.

Manyika, J., Lund, S., Bughin, J., Robinson, K., Mischke, J. and Mahajan, D. (2016). *Independent Work: Choice, Necessity, and the Gig Economy*. McKinsey Global Institute Report, McKinsey and Co.

Maslen, H., Savulescu, J. and Hunt, C. (2020) “Praiseworthiness and Motivational Enhancement: ‘No Pain, No Praise’?,” *Australasian Journal of Philosophy*, 98(2): 304-318

Metz, T. (2013). *Meaning in Life*. Oxford: Oxford University Press.

Nyholm, S. (2018). Attributing agency to automated systems: Reflections on human-robot collaborations and responsibility-loci. *Science and Engineering Ethics*, 24(4), 1201–1219.

Nyholm, S. (2020). *Humans and robots: Ethics, agency, and anthropomorphism*. London: Rowman & Littlefield

Nyholm, S., Smids, J. Can a Robot Be a Good Colleague?. *Sci Eng Ethics* (2019). <https://doi.org/10.1007/s11948-019-00172-6>

Prassl, Jeremias (2019). *Humans as a Service: The Promise and Perils of Work in the Gig Economy*. Oxford, UK: Oxford University Press.

Rosenblat, Alex (2019). *Uberland: How Algorithms are Re-writing the Rules of Work*. Oakland, CA: University of California Press.

Smids, J., Nyholm, S. & Berkers, H. Robots in the Workplace: a Threat to—or Opportunity for—Meaningful Work?. *Philos. Technol.* 33(3) (2020): 503-522. <https://doi.org/10.1007/s13347-019-00377-4>

Smith, Adam (1776) *The Wealth of Nations*. London: W. Strahan and T. Cadell.

Smuts, A. (2013). The good cause account of meaning in life. *Southern Journal of Philosophy*, 51(4), 536–562.

Susskind, Daniel (2020). *A World Without Work*. London: Penguin.

Taplin, J. (2017) *Move Fast and Break Things*. PanMacMillan

Tigard, D. (2020a). Artificial moral responsibility: How we can and cannot hold machines responsible. *Cambridge Quarterly of Healthcare Ethics*, forthcoming

Tigard, D. (2020b). There is no Techno-Responsibility Gap. *Philosophy and Technology*, forthcoming available at: <https://doi.org/10.1007/s13347-020-00414-7>

Veltman, A. 2016. *Meaningful work*. Oxford: Oxford University Press

Weil, A. (2014). *The Fissured Workplace: How Work Became So Bad for So Many and What Can be Done About It*. Harvard: Harvard University Press

Wolf, S. (2010). *Meaning in Life and Why it Matters*. Princeton: Princeton University Press