Chapter 4
Ethical AI at Work: The Social Contract for Artificial Intelligence and Its Implications for the Workplace Psychological Contract

Sarah Bankins and Paul Formosa

4.1 Introduction

The current fourth industrial revolution is significantly disrupting the world of work (World Economic Forum, 2018). One driver of this disruption is the increasing use of artificially intelligent (AI) technologies in workplaces. As these technologies change how work tasks are completed and which tasks are done solely by humans, which are done solely by AI technologies, and which are completed by both in collaboration, this will alter how employees view their employment relationships. Examining the psychological contract (PC) and what shapes it in such contexts offers one way to assess the implications of AI technologies for the employee–employer relationship. The PC constitutes "a cognitive schema, or system of beliefs, representing an individual's perceptions of his or her own and another's obligations, defined as the duties or responsibilities one feels bound to perform" (Rousseau, Hansen, & Tomprou, 2018, p. 1081). Individual perceptions of PC fulfilment (met obligations) or PC breach (unmet obligations) generally lead to positive employee responses in the case of the former (Parzefall & Hakanen, 2010) and negative employee responses in the case of the latter (Zhao, Wayne, Glibkowski, & Jesus, 2007). Overall, understanding the factors that shape the formation of the PC, its subsequent content, and its degree of fulfilment can help to explain the attitudes and actions of employees at work (Conway & Briner, 2005).

S. Bankins (✉)
Department of Management, Macquarie Business School, Macquarie University, Sydney, NSW, Australia
e-mail: sarah.bankins@mq.edu.au

P. Formosa
Department of Philosophy, Macquarie University, Sydney, NSW, Australia
e-mail: paul.formosa@mq.edu.au

M. Coetzee, A. Deas (eds.), *Redefining the Psychological Contract in the Digital Era*, https://doi.org/10.1007/978-3-030-63864-1_4

As the nature of work has changed over time, such as through economic and labour market deregulation, researchers have utilised the PC to examine the impact of such disruptions on the employment relationship (Schalk & Rousseau, 2008). However, despite a new raft of changes being driven by the expanding role of AI at work, PC research is yet to widely engage in examining the implications of these changes, or arguably the role of technology more broadly, for the employment exchange. This is despite evidence that intelligent technologies such as social robots (Bankins & Formosa, 2020) and smartphones (Obushenkova, Plester, & Haworth, 2018) will influence the employee–employer obligations underpinning the PC and may even drive technology itself to be viewed by employees as a contracting partner (or counterparty) to the PC (Bankins, Griep, & Hansen, 2020). In particular, employers' decisions regarding how they implement AI alongside, or in replacement of, their employees are increasingly recognised as having ethical dimensions. This generates sets of obligations upon employers to deal with such decisions appropriately. In this theoretical chapter, we offer one pathway for addressing this 'technology gap' in the PC literature.

4.2 Chapter Objective

Our chapter focuses on examining the growing range of factors that are encouraging an ethical application of AI at work to then argue, and demonstrate how, these factors are likely critical inputs into individuals' PC content. In outlining these factors, operating at multiple levels, we suggest how they will shape organisational implementation of AI, how they will influence groups' and individuals' views of AI in the workplace, and how they will feed into employees' workplace PCs. Specifically, we utilise Integrative Social Contracts Theory (ISCT) to outline the key multilevel influences shaping the ethical use of AI in the workplace, as well as the notion of technology frames at the organisational and individual levels, and their links to the PC. We begin the chapter by briefly outlining the nature of AI and how it is being used in workplaces. We then outline ISCT and detail what the macrosocial and various microsocial contracts for the ethical creation and use of AI at work look like. Finally, we explain how these various social contracts act as a normative background that will inform the individual PCs of workers, before demonstrating our account through an illustrative example. We conclude by outlining theoretical and practical implications of our work.

4.3 Artificial Intelligence at Work

Artificial intelligence refers to "the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings (i.e. humans) … such as the ability to reason, discover meaning, generalise, or learn

from past experience" (Copeland, 2020). Functionally, AI refers to technology that does intelligent things, such as tasks that would require the use of intelligence were a human to perform them (Boden, 2016; Floridi & Cowls, 2019; Walsh et al., 2019). Machine learning through supervised, semi-supervised, or unsupervised training is a common feature of many AI systems. This training allows AIs to perform a range of tasks, such as natural language processing (speech recognition and production), image recognition and classification (including facial recognition), and goal-directed reasoning and decision-making activities such as planning, scheduling, and optimising the use of resources (Walsh et al., 2019). All current AI applications are examples of different forms of artificial *narrow* intelligence, which can perform intelligent functions only within restricted domains and which lack the ability to quickly transfer learned skills to other domains. For example, Deepmind's *AlphaGo* is expert at the game GO, but it cannot hold a conversation or recognise cats (Robbins, 2019). In contrast, artificial *general* intelligence refers to an AI that can perform at a similar level to humans across all intelligent activities (Bostrom, 2014). However, given broad disagreement about how imminent artificial general intelligence is (Bostrom, 2014), we restrict our focus to artificial narrow intelligence as a technology that is already widely used and being continually improved (Boden, 2016).

Currently, AI is predominantly used in workplaces to automate specific tasks rather than to replace whole occupations, except where jobs, often in manufacturing, involve simple and repetitive tasks that can be wholly automated (Walsh et al., 2019). The varied skill sets of narrow AI are already widely utilised across many industries (Bekey, 2012; Walsh et al., 2019). For example, in the service sector AI is deployed at customer-facing points of contact in the form of chatbots and virtual assistants. The optimisation abilities of AI are used across farm and mine management, logistics, resource allocation, and in military contexts to manage resources effectively (Walsh et al., 2019). Artificial intelligence also powers autonomous vehicles for use in transportation, mining, farming, and manufacturing. In healthcare, AI can support diagnoses, generate health insights, and offer personalised patient care by drawing on large data sets (Walsh et al., 2019), while in the legal and criminal justice sectors AI technologies help locate legal precedents and advise on parole and sentencing decisions (Angwin, 2016). In defence and security contexts, AI supports intelligence collection and analysis (including the use of facial recognition), cybersecurity operations, and autonomous weapons systems (Sharkey, 2013). Artificial intelligence use is also prevalent in the FinTech sector for fraud detection, risk management, compliance checks, and for intelligent share trading agents (Wellman & Rajan, 2017). It is also increasingly used in human resource management for recruitment and selection (Albert, 2019) and work allocation (Lee, Kusbit, Metsky, & Dabbish, 2015).

The use of AI in workplaces differs from other technologies in two key ways. First, the capabilities of AI technologies are already impressive, as they can surpass human capabilities in tasks such as synthesising and analysing data and generating predictions across large data sets (see Walsh et al., 2019). This has extended the scope of work that AI can undertake into areas that traditionally required human cognition (Copeland, 2020). Second, and relatedly, AI capabilities are now shifting

the mix of what organisations understand to be human work and machine work (World Economic Forum, 2018), thereby changing the type and amount of work humans do. For example, by 2030 it is forecast that up to 375 million workers will have switched occupational categories as a result of automation (Yaxley, 2019) and by 2022 over half of all employees will require significant skill changes due to the use of AI at work (World Economic Forum, 2018).

Taken together, this implies that AI will significantly change what workers do in their jobs, how they experience their employment, and what they understand to be reciprocal employee–employer obligations (i.e. their PCs). The potential scale of AI's impacts on human workers also raises ethical questions regarding its application, such as for what purposes it is used in workplaces. That is, while the use of AI offers many potential benefits to workers, such as improving decision-making and talent management (e.g. Colson, 2019; Guenole & Feinzig, 2018), its deployment can also generate harms. For example, biased data feeding machine learning recruitment algorithms can negatively discriminate against job applicants based on gender or race (Tambe, Cappelli, & Yakubovich, 2019), the design of algorithms may fail to deliver fair outcomes across different contexts (Selbst, boyd, Friedler, Venkatasubramanian, & Vertesi, 2019), and algorithmic management may exert new forms of control over workers (Kellogg, Valentine, & Christin, 2020) that ultimately restricts their autonomy. To examine the impact of AI use on the employment relationship through an ethical lens, we draw on ISCT (Donaldson & Dunfee, 1994, 1995) and technology frames (Orlikowski & Gash, 1994) to develop a model of the cascading effects of multiple forces, or different forms of contracts, that will shape the individual-level PC in workplaces that are deploying AI technologies.

4.4 Overview of Integrative Social Contracts Theory

Integrative Social Contracts Theory (ISCT) has been used extensively in the field of business ethics to examine the ethical implications of organisational actions and has been applied to many types of decisions, such as downsizing (Van Buren, 2000). It is a social contracts theory, positing the existence of macrosocial and microsocial contracts. Applied organisationally, the terms of these contracts bind the behaviours of actors, such as leaders, and guide them when making decisions with ethical implications.

The macrosocial contract is a hypothetical social contract comprised of "the set of principles regarding economic morality to which [rational] contractors would agree [to]" under conditions where universal consensus by all affected persons is required (Donaldson & Dunfee, 1994, p. 260). The macrosocial contract sets the "hypernorms", which are likened to universal human rights, that create an "ethical floor" for subsequent microsocial contracts (Van Buren, 2000, p. 210). In the domain of employment, authors have drawn on the work of international organisations, such as the United Nations, the Organization for Economic Cooperation and Development (OECD), and the International Labour Organization, to suggest a macrosocial

contract focused on organisations providing, among other things, meaningful work, "a living wage and a healthy and fair work environment" (Wright & Schultz, 2018, p. 5).

The microsocial contract is a context-specific and community-based social contract that is limited and informed by the underlying macrosocial contract. "Community" is understood here to refer to "a self-defined, self-circumscribed group of people who interact in the context of shared tasks, values, or goals and who are capable of establishing norms of ethical behaviour for themselves" (Donaldson & Dunfee, 1994, p. 262), such as organisations (Smith, 2000). Within the boundaries set by the macrosocial contract's hypernorms, different communities may develop their own sets of detailed context-specific norms through microsocial contracts which reflect the ethical life of these communities and the preferences of their members (Donaldson & Dunfee, 1999). However, such norms must be grounded in informed consent (individuals can meaningfully agree to them) and a right of exit (individuals can remove themselves from that contract). Where various microsocial contracts exist and their norms conflict with one another, ISCT offers six rules of thumb to help resolve such conflicts, such as giving priority to norms that are grounded in larger communities, norms that are consistent with other norms, and norms that are well defined (Donaldson & Dunfee, 1994). In a workplace context, the microsocial contract can be likened to what the PC literature terms a normative contract, which develops where groups of employees form a shared understanding of the obligations their organisation has towards them as a group (Rousseau, 1995). The content of the normative contract has been shown to influence how employees develop and evaluate their individual PCs (Cregan, Kulik, Metz, & Brown, 2019; Estreder, Rigotti, Tomás, & Ramos, 2020).

The content of the macrosocial and microsocial contracts guides the ethical behaviour of agents in a community by shaping their degree of "moral free space" (Dunfee, 2006, p. 315). That is, macrosocial and microsocial contract norms provide rules by which members must abide, and to which they are held accountable by stakeholders, when taking decisions and actions. This constrains the types of decisions and actions actors can take. "Unoccupied moral free space" exists when there are no clear hypernorms (at the macrosocial level) or legitimate other norms (at the microsocial level) that can be applied to the decision at hand, meaning the decision maker must then rely on personal views or values (Dunfee, 2006, p. 315). The norms constituting the microsocial contract are needed to help overcome the vagueness and generality of hypernorms and to allow communities to fill in some of the moral free space that those hypernorms create (Donaldson & Dunfee, 1994).

The idea that the PC is embedded within these wider contracting processes has gained currency over time (e.g. Thompson & Hart, 2006; Van Buren, 2000). That is, while the PC, as an individual-level construct, sits below the macrosocial and microsocial contracts, PCs will be "strongly influenced by … norms" within each of these sets of contracts (Thompson & Hart, 2006, p. 239). This is because these universal and community-based normative contracts will help inform the perceived obligations employees hold regarding their individual treatment, which will in turn constitute the content of their individual PCs. Conversely, PCs can feed back into, and

help to re-define, microsocial contracts through their impacts on the everyday lived experiences of individuals (Thompson & Hart, 2006). Overall, it is recognised that both higher-level contracts (macrosocial and microsocial) will inform the content of individuals' PCs. We now turn to applying the ISCT framework to the context of increasing AI use at work.

4.5 Applying ISCT to the Ethical Use of AI at Work: From Macrosocial to Microsocial to Psychological Contracts

In the following sections we sketch the emerging content of the higher-level contracts (macro- and microsocial) regarding the ethical use of AI that we argue will input into, and help us to understand, the PCs of employees whose workplaces are increasingly integrating AI technologies.

4.5.1 Macrosocial Contract

We begin our framework by drawing on recent work (e.g. Floridi et al., 2018; Winfield, Katina, Pitt, & Evers, 2019) formulating broad principles, or the highest level set of basic norms, for ethical AI to sketch an emerging macrosocial contract for the ethical use of AI *in workplaces*. Ethical AI refers to the fair and just development, use, and management of AI technologies. While many competing guidelines for ethical AI have been developed in several countries, leading to concerns about "principle proliferation" (Floridi & Cowls, 2019, p. 2), several recent literature reviews have systematised and grouped these principles (Floridi & Cowls, 2019; Hagendorff, 2020; Jobin, Ienca, & Vayena, 2019). The strong overlaps between these reviews suggest that a global consensus regarding ethical AI principles is emerging, providing a promising basis for a macrosocial contract for AI (see also Rahwan, 2018, for a related discussion).

Floridi and Cowls (2019) argue for five ethical AI principles: beneficence; non-maleficence; autonomy; justice; and explicability (cf. Floridi et al., 2018). According to these principles AI should: benefit people, promote well-being, and be environmentally sustainable (beneficence); respect privacy and not harm people (non-maleficence); allow people the power to decide what to do where possible (autonomy); be fair and equitable, avoid bias, and preserve solidarity (justice); and its decisions should be intelligible to us and accountability for its decisions should be clear (explicability). Jobin et al. (2019) thematically analyse various international ethical AI documents and argue that these contain 11 overarching ethical values and principles which are, in order of frequency: transparency; justice and fairness; non-maleficence; responsibility; privacy; beneficence; freedom and

autonomy; trust; dignity; sustainability; and solidarity (for a similar review see Hagendorff, 2020). Although these reviews group principles in different ways, they significantly overlap. For example, Jobin et al.'s separate principles of "solidarity" and "justice and fairness" are together grouped under Floridi and Cowls' broader category of "justice". Given their greater simplicity and generality, we use Floridi and Cowls' (2019) five principles for ethical AI in our framework.

From an ISCT perspective, we argue that the emerging consistency of such hypernorms serves to constrain the moral free space of organisational leaders, even if only thinly (Smith, 2000), regarding their implementation of AI in workplaces, and therefore what employees believe their organisation's obligations are in this regard. At a macrosocial level, it is reasonable to position these hypernorms as broadly and universally endorsed, and therefore organisations will be bound to implement AI within them. However, the broadness of hypernorms leaves plenty of moral free space which must be filled in via norms and obligations at the microsocial contract and, ultimately, PC levels. For example, what counts as "environmentally sustainable" within the "beneficence" hypernorm needs further specification, and indeed may vary, in different communities. Such clarification of the macrosocial contract for the ethical use of AI can occur through codes, policies, and practices designed to specify authentic implementations of these background ethical principles in discrete contexts.

4.5.2 Microsocial Contracts

When microsocial contracts operationalise hypernorms within community contexts, they also significantly shape the moral free space of the actors within them. Therefore, the microsocial level is a key site for understanding moral free space as it relates to shaping organisations' decisions and actions towards the ethical use of AI, which will influence how employees' experience AI at work. The scope of potential communities that develop the concrete norms at this level is large, ranging from the national level to organisational teams. We argue that microsocial contracts will likely develop at three key levels, although these may overlap or merge in some cases: national; industry; and organisational.

Microsocial contract: National-level norms. Many countries, and in some cases regional blocs such as the European Union, have developed norms through guidelines and discussion papers focused on the ethical use of AI, including in the workplace. For example, Hagendorff (2020) contrasts the privacy principles of national- and regional-level ethical AI documents from the United States of America, the European Union, China, and the OECD. While these largely overlap and cluster around the five principles identified above at the macrosocial level, they nonetheless differ significantly in length and technical detail (ranging from 22,787 words to 766 words) and emphasis (e.g. privacy is a key issue in the European Union document but is less important in the US document) (Hagendorff, 2020). National and regional governmental initiatives can place further constraints on how AI is used by

organisations in those nations, thus generating a community-specific (national-level) microsocial contract. This will either further constrain or leave open, beyond the macrosocial contract, organisations' moral free space regarding AI adoption within workplaces. However, some authors argue that governmental regulations will be targeted towards AI implementation in specific sectors (Dasgupta & Wendler, 2019), meaning that industries can also constitute an important community for microsocial contract development.

Microsocial contract: Industry-level norms. Where industries or sectors have been early adopters of AI and automation, we suggest that this will likely, as the below examples show, generate industry norms that will further shape the moral free space for individual organisations' implementation of AI. These industry norms can be expressed through formal documents that organisations explicitly sign up to, or informal and de facto norms that arise implicitly. An example of the former is the *Partnership on AI* which has over 100 partners across 13 countries, including global technology leaders, such as Amazon, Apple, Facebook, Google, and Microsoft, non-profit organisations, such as Amnesty International, and university and media organisations, such as *The New York Times*. Partner organisations endeavour to uphold eight ethical AI tenets, such as using AI to "benefit and empower" people (aligned to the beneficence hypernorm), being "accountable to a broad range of stakeholders" (aligned to the explicability hypernorm), and protecting the privacy and security of individuals (aligned to the non-maleficence hypernorm) (Partnership on AI, n.d.). This creates pressure on partner organisations to be seen to be abiding by these tenets.

An example of informal industry norms can be found in the mining sector, which is recognised as an early adopter industry for AI and automation, where ethical AI principles appear to have emerged informally through practice. The nature of some types of mining work means it can be occupationally hazardous, with the industry accounting for up to 5% of workplace fatalities globally (or roughly 15,000 deaths per year) (Amin, 2018). As a result, it could be argued that AI and robotic technologies have been implemented with a focus on automating "dull, dirty, and dangerous" work and work in highly remote locations (Marr, 2017). For example, at mine sites in the Pilbara in Western Australia, fully autonomous (or robotic) vehicles conduct activities such as haulage, an activity that has historically caused significant workplace injuries (Amin, 2018; Marr, 2017). While the application of automation in this way is likely designed with efficiency in mind, it also increases worker safety and limits the extent to which dangerous work is undertaken by humans (Amin, 2018). This is an example of operationalising, in an industry context, the macrosocial contract hypernorms of beneficence and non-maleficence through using AI to protect workers' safety and well-being. Therefore, it could be argued that in sectors such as this, with established or emerging patterns of workplace AI use, general industry norms develop that guide how other organisations in that sector adopt the same technologies. Indeed, studies of information system adoption more broadly note such isomorphic pressures that organisations face (Pal & Ojha, 2017). Whatever way such norms emerge, explicitly or implicitly, we suggest they generate a

community-level microsocial contract that either constrains or leaves open an organisation's moral free space for AI implementation.

Microsocial contract: Organisational-level norms. The individual organisation is a key location for microsocial contract development. Just as at the industry level, these norms can form either explicitly or implicitly. An example of explicit organisational-level norms are Microsoft's (n.d.) six AI principles, which include "Fairness" and "Inclusiveness" (aligned to the justice hypernorm), and "Transparency" and "Accountability" (aligned to the explicability hypernorm). The associated videos explain where and how, at an organisational level, these principles are implemented. For example, the organisation shows how the Accountability, Fairness, and Privacy and Security principles limit how Microsoft develops and sells facial recognition technology, and how the Inclusiveness principle requires that speech recognition software is trained to work for minority groups (Microsoft, n.d.). Drawing on our ISCT framework, we can see how these explicit organisational-level microsocial norms in turn sit under relevant industry norms (the industry-level microsocial contract), such as those expressed in the *Partnership on AI*'s tenets of which Microsoft is a partner, which in turn (where applicable) sit under relevant national documents (the national-level microsocial contract), as well as ultimately under the universal macrosocial contract's hypernorms.

Organisational norms regarding ethical AI use can also be developed implicitly, and we suggest that group-level technology frames can help to create informal organisational norms regarding AI use at work. Technology frames refer to one's assumptions and knowledge about technology and its uses (which can become shared at a group level), and these beliefs shape how people make sense of and respond to technology in organisations (Orlikowski & Gash, 1994). As such, technology frames can be conceptualised at both the group and individual levels. Such frames can centre on views about: technology-in-use (views on how technology is used on a daily basis and the conditions and consequences of use); technology strategy (views about why the organisation implements certain technologies and the motivation for and vision behind implementation); and technology nature (general views of a technology and its capabilities and functionality) (Orlikowski & Gash, 1994). Such frames help explain how and why employees respond to certain technologies and help to capture the day-to-day experience of enacting explicit organisational technology-related norms. This will likely inform group-level norms about AI use within organisations. For example, employees' beliefs regarding the appropriate use of AI ("technology-in-use" frame) could translate into the belief that their employer is obligated to only implement AI for some tasks and not others, thus generating relevant norms at the organisational level. Wright and Schultz (2018) also offer examples of workers collectively voicing dissent towards the use of robots, automation, and algorithmic management in their workplaces (again related to the "technology-in-use" frame).

While the content of these informal organisational norms, driven by group-level technology frames, could potentially be wide and different across organisations, we suggest that such norms will cluster around three broad categories: *AI receptive*; *AI neutral* (indifferent); and *AI resistant*. For example, an informal norm may develop

amongst employees that they are "receptive" to the automation of monotonous tasks, but "resistant" to having AI undertake customer-facing tasks that are viewed as requiring human involvement. This means that each organisation will likely have different combinations of informal norms around receptivity, resistance, and neutrality to AI being integrated into various aspects of work. However, it is up to the organisational community in the microsocial domain (or potentially multiple communities within it) to set such norms. Taken together, we suggest that the nature of these norms will shape the moral free space of organisations and their leaders regarding how and in what ways they use AI.

We do, however, place a caveat on our arguments. The extent to which group-level technology frames exist and drive informal organisational norms can depend on the homogeneity of employee groups in an organisation. That is, the extent to which individuals in the organisation do similar forms of work and have similar organisational experiences. Research on shared team perceptions of PC fulfilment suggests that such perceptions are more likely to develop when members experience similar events and regularly interact and share information (Laulié & Tekleab, 2016). For example, in workplaces largely staffed by high-skilled, white collar professionals undertaking similar work, consistent group-level technology frames are more likely to emerge than in workplaces where diversely skilled workers undertake very different types of work. Amazon provides one example of such a diversified organisation, as the technology (i.e. robotics) deployed in warehouses alongside lower-skilled, blue collar workers is not similarly deployed alongside head office workers (e.g. Sainato, 2020). Overall, this means that where there is higher homogeneity in work, and thus similar experiences across employee groups, there are likely to be more consistent group-level technology frames and so more likely to be a singular microsocial contract regarding norms for AI use. However, where there is less homogeneity (e.g. Amazon), there are likely to be more fragmented group-level technology frames and so multiple, potentially conflicting, microsocial contracts regarding norms for AI use. It has also been found that differences in technology frames can exist across different organisational groups, such as managers and lower-level employees (Orlikowski & Gash, 1994). To resolve such conflicts, as identified earlier, ISCT stipulates rules of thumb to help prioritise any conflicting norms in different microsocial contracts.

4.6 Psychological Contracts

By providing a normative background, the relevant macrosocial and microsocial contracts are important inputs into the content of individuals' PCs and their evaluation of them (Thompson & Hart, 2006; Van Buren, 2000). In this section, we focus on exemplifying how these higher-level norms may cascade into the individual-level PC. For example, the macrosocial hypernorm of explicability (i.e. AI decisions must be intelligible and accountable) might be contextualised within an industry that commonly uses AI algorithms to determine employees' bonuses and

promotions to mean that workers can request a timely human review of these algorithmic determinations (industry-level macrosocial contract). This in turn could be further clarified through, for example, organisational-level microsocial norms to mean that workers may appeal to their line manager within 30 days to have an algorithmic decision reviewed and explained. This normative background will likely then shape the PC (individual level) obligations that employees perceive between themselves and their employer about, in this case, the explanation and review process they can expect when subject to an algorithmic decision. When these PC beliefs are not met, such as when an inadequate explanation for an algorithmic decision is given or no recourse to human review is provided, then a PC breach will likely be perceived from which negative employee responses may ensue (Morrison & Robinson, 1997). More generally, we suggest that the various macrosocial and microsocial contracts around ethical AI will help to drive the technology-specific components of individuals' PCs, such as perceived obligations regarding re- and up-skilling in the use of AI technologies, expectations for supporting workers displaced by AI, and future organisational plans for AI implementation across individual tasks and roles.

However, individual uptake and incorporation of these macrosocial and microsocial norms into individual PCs will likely vary depending upon individual endorsement of those norms. As identified earlier, the strength of group-level technology frames, which we argue are important elements of the organisational-level microsocial contract for ethical AI use, may vary (Treem, Dailey, Pierce, & Leonardi, 2015) or may not exist at all when individuals have highly dispersed views of the role of AI in the workplace. Where group-level technology frames are strongly held and generate clear organisational-level microsocial norms, we suggest those frames will feed into individuals' PCs. However, where group-level technology frames are inconsistent or dispersed, we suggest that individuals will likely develop more individualised views of the role of AI in the workplace, through individual-level technology frames, which will inform their PCs. Because the exact content of any particular PC in regard to the use of AI technologies will depend on the individual and their experiences, their industry, their organisation, and the nature of their work and interactions with AI (see Rousseau (1995) for the range of potential PC inputs), we now offer an illustrative example to demonstrate how our framework may unfold in practice.

4.7 Illustrative Example: Telenor

This example aims to show how the different levels of contracts we have derived through an ISCT-based assessment of ethical AI use in the workplace will influence the individual-level PC. It should be noted that the example is not intended to be exhaustive as the data used to generate it are derived from secondary sources, particularly the focal company's website and other publicly available communications.

The example company is Telenor, a telecommunications company based in Norway (see Telenor Group, 2019a).

Because ISCT positions the macrosocial contract as universal, we do not outline it again here, but instead take for granted the five ethical AI principles outlined earlier. At a microsocial level, we suggest that multiple communities exist that are relevant to Telenor's implementation of AI and thus the PCs of its workers. At a national level, Norway has developed a strategy for the development and implementation of AI through its "National Strategy for Artificial Intelligence" launched in January 2020 (NORA, 2020). Interestingly, and aligned with ISCT, the Norwegian government acknowledges that such a strategy cannot, and does not aim to, guide every aspect of AI development and use in Norway, but instead "will give a direction and thus serve as a framework for both public and private entities seeking to develop and use Artificial Intelligence" (NORA, 2020).

Such a national strategy demonstrates alignment with the wider macrosocial contract for AI, by explicitly identifying the need to respect "ethical principles" and "human rights", safeguard individual privacy, and operate in accordance with "the principles for responsible and trustworthy use of Artificial Intelligence" (NORA, 2020). The strategy also notes that the development and use of AI is critical for organisations' ongoing competitive advantage, thus leaving open some moral free space for specific organisational norms for the uptake of AI. The national strategy also supports widening educational opportunities for upskilling in the understanding and use of AI, for example through the #AIchallenge which (in part) tasks large companies with supporting employees in completing an online AI course (Telenor Group, 2020). Such work is supported within the European Union through the bloc's efforts to develop a pan-European approach to supporting ethical AI development and use, including through cross-organisational and cross-country data sharing (European Commission, 2020). Telenor is thus situated in a national and regional bloc that is active in developing explicit regulations and guidance for the development and ethical implementation of AI for national advantage. This constitutes the regional- and national-level microsocial contracts within which Telenor operates.

At an organisational level, a microsocial contract for Telenor itself can be sketched (although we acknowledge its limitations as it is solely based on the company's public communications). Telenor aims to become "a data-driven company, where AI/ML (machine learning) capabilities will be an asset, and … create a competitive advantage" (Telenor Group, n.d.). The company is explicit regarding where it will target the use of AI, such as for "optimizing network operations, automating customer interactions, personalizing marketing and sales campaigns", as well as areas of ambition in the deployment of AI such as strengthening existing data analytic capabilities and extending into new areas like IoT (Internet of Things) (Telenor Group, n.d.). The norms within this community-level contract also extend to fostering collaborations with academia and public and private sectors (creating a "*dugnad*" or joint work for collective benefit) to strengthen the development of AI capabilities in the Norwegian population, including Telenor employees (Telenor Group, 2020). Telenor is also a signatory to, and a leading developer of, a Norwegian "Declaration for the Responsible Use of AI in Working Life", which is an

industry-level initiative to embed ethical principles for the use of AI into workplaces, including ensuring that "AI-supported decisions and recommendations (are) fair, non-discriminatory and transparent" and that data privacy is protected (Telenor Group, 2019a). While, given the nature of the data we are using, we refrain from articulating any group-level technology frames, we do suggest that the organisational-level microsocial contract at Telenor seems to be generally *AI receptive* across a range of the technology's potential deployments.

Aligned to our framework, we suggest that each of these contracts will then input into individual employees' PCs. We focus particularly on the idea of AI upskilling, which feeds through from the national- and organisational-level microsocial contracts. The expectation for employees to upskill is clear through organisational initiatives, such as adding AI competencies to the Telenor Campus (an employee training initiative), Telenor's "40-hour challenge" where employees can access 40 hours of training to upskill in AI (Telenor Group, 2020), and a belief that "lifelong learning (is) the new normal" (Telenor Group, 2019b). This suggests that ongoing and lifelong employee skill development regarding AI appears to be important within the employment exchange. More broadly, at the individual level of the PC, we suggest that reciprocal employee–employer obligations will be largely *AI receptive* and likely centre on beliefs around: expected privacy of one's data when it is used in AI applications by the company (aligned to the non-maleficence hypernorm); expectations around the skills, and ongoing up-skilling, required to understand and use AI (aligned to the beneficence hypernorm); the specific areas of the workplace in which AI will be deployed (aligned to the justice hypernorm); the explainability of AI's decisions when they impact workers (aligned to the explicability hypernorm); and that AI implementations will be respectful of employees' rights and freedoms, such as levels of autonomy, when it is used (aligned to the autonomy hypernorm).

4.8 Implications for Theory

We offer here several theoretical extensions for the PC literature. First, the fourth industrial revolution is driving the integration of increasingly sophisticated forms of workplace technologies. These already, and promise to increasingly do so in the future, shape employees' work experiences and the technology-specific components of their PCs. However, PC literature is lagging in its examinations of these impacts. By centrally positioning AI technology and its use as a key driver of PC beliefs, we show how its influence will increasingly shape the nature of the employee–employer exchange. To this end, we also demonstrate the utility of integrating technology-specific frameworks into the study of PCs, particularly by outlining the role of technology frames in driving norms of AI resistance, neutrality, or receptivity amongst employees, which we argue will flow through to beliefs about employer–employee obligations.

Second, early PC theorising (e.g. Rousseau, 1995) recognises that the content of these contracts will develop from many sources, including extra-organisational sources. Recent contemporary work reinforces this with, for example, Rousseau et al. (2018) suggesting that normative expectations, often derived from sources beyond the organisation, will be important inputs into the PC. However, PC research is yet to clearly articulate what these sources may be in a given context. Without this specificity it can be difficult for researchers to know which sources of information may be most influential for informing the PC and so which to focus their analyses upon. In the context of increasing use of AI technologies in the workplace, and by utilising ISCT, we specify those sources of information at various levels and we argue that these will function as inputs into individual-level PCs. Indeed, when AI is newly implemented into organisations, it may primarily be global macrosocial and national- and industry-level microsocial contracts that inform employees' technology frames and PCs. Overall, our framework highlights the need to embed PC research in the multilevel contexts in which it is situated, particularly when analysing the impacts of a global phenomenon such as AI.

Third, while the PC is largely studied as an individual-level construct, its shared nature is increasingly recognised (e.g. Laulié & Tekleab, 2016). Our work helps extend this line of investigation by highlighting the need to investigate multiple microsocial contracts amongst diverse employee groups and, when technology in the workplace is the focal context, to examine the role of group-level technology frames as a form of shared cognitions that will also inform individuals' PCs. As the microsocial level is where multiple and potentially competing contracts may emerge, future PC work could also examine how groups of employees compare and contrast their different microsocial contracts and how this informs perceptions of breach at both group and individual levels.

As a theoretical chapter, our work comes with limitations. While we have sought to ground our use of ISCT with evidence-based examples, the actual content of the norms within the macrosocial and microsocial contracts can only be determined through empirical work. Such norms, as we identify, are also likely to differ across countries, sectors, and organisations, meaning our framework is designedly broad to capture this contextual diversity, but with trade-offs regarding specificity to any one context. Also, while we have argued that macrosocial and microsocial contract norms will input into the PC (supported by others, such as Thompson & Hart, 2006), it becomes an empirical question as to whether employees are actually cognizant of, and therefore draw upon, these higher-level norms in ultimately formulating and evaluating their PCs.

4.9 Implications for Organisational Practice

Several organisational implications also stem from our work. For example, the evidence we have drawn on suggests that macrosocial and microsocial contracts for ethical AI, particularly at the national and industry levels, remain quite "thin". This

means that the moral free space for many organisations and their leaders regarding how they operationalise these norms within these contracts remains comparatively large. Practically, such thinness means companies can have little guidance on the ethically optimal ways to implement AI in their workplaces. The macrosocial hypernorms of justice or beneficence, for example, are intentionally broad and do not, in isolation, fully determine where and how an organisation should use AI or, notwithstanding ISCT's rules of thumb, suggest how best to deal with any conflicts between hypernorms. For example, a company may seek to tailor employee training opportunities by utilising sensitive employee performance and historical training attendance data, thus increasing employees' benefits (beneficence hypernorm) but at the cost of lowering their privacy protections (non-maleficence hypernorm). Thus, when higher-level contracts are left fairly thin this leaves scope for potentially positive, but also potentially more detrimental, applications of AI in the workplace. It may be that national- and industry-level microsocial contracts are particularly important to have in place early in the adoption of AI across sectors to help inform organisational microsocial contracts and ultimately shape employees' PCs.

Further, as organisations implement AI in their workplaces, leaders should be mindful that employees' pre-existing views of this technology, both individually and collectively, will shape where and how they believe AI should be adopted and how its use will alter reciprocal employee–employer obligations. In understanding employees' responses to AI use, it will be beneficial for leaders to identify what group-level technology frames may be operating and whether those norms are clustering around AI receptivity, neutrality, or resistance. They could then seek to either influence those frames, or shape AI implementation to align with them, to minimise the potential for individual-level PC breaches and the negative consequences that often flow from these.

4.10 Conclusions

Overall, we seek to contribute to a growing body of literature recognising that the PC is developed within an organisational, national, regional, and global context. With a focus on the increasing use of AI in the workplace, we argue that the macrosocial and microsocial contracts for the ethical use of AI will be important inputs to the PC, and we outline the factors contributing to such higher-level contracts. From a practical perspective, our work should inform both policymakers and organisational leaders regarding how higher-order sets of norms will ultimately influence the micro-level reciprocal obligations between employees and employers (the PC).

References

Albert, E. T. (2019). AI in talent acquisition: A review of AI-applications used in recruitment and selection. *Strategic HR Review, 18*(5), 215–221.

Amin, C. (2018, December 10). *Mining industry automation: "Let the robots take our dangerous and dirty jobs"*. Retrieved from https://www.createdigital.org.au/mining-industry-automation-robots/

Angwin, J. L. (2016, May 23). *Machine bias*. ProPublica. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bankins, S., & Formosa, P. (2020). When AI meets PC: Exploring the implications of workplace social robots and a human-robot psychological contract. *European Journal of Work and Organizational Psychology, 29*(2), 215–229.

Bankins, S., Griep, Y., & Hansen, S. (2020). Charting directions for a new research era: Addressing gaps and advancing scholarship in the study of psychological contracts. *European Journal of Work and Organizational Psychology, 29*(2), 159–163.

Bekey, G. A. (2012). Current trends in robotics: Technology and ethics. In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot ethics* (pp. 17–34). Cambridge: MIT Press.

Boden, M. A. (2016). *AI: Its nature and future*. Oxford: Oxford University Press.

Bostrom, N. (2014). *Superintelligence*. Oxford: Oxford University Press.

Colson, E. (2019). *What AI-driven decision making looks like*. Harvard Business Review. Retrieved from https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like

Conway, N., & Briner, R. B. (2005). *Understanding psychological contracts at work*. Oxford: Oxford University Press.

Copeland, B. J. (2020, March 24). *Artificial intelligence*. Encyclopedia Britannica. Retrieved from https://www.britannica.com/technology/artificial-intelligence

Cregan, C., Kulik, C. T., Metz, I., & Brown, M. (2019). Benefit of the doubt: The buffering influence of normative contracts on the breach-workplace performance relationship. *The International Journal of Human Resource Management*. https://doi.org/10.1080/09585192.2018.1528471

Dasgupta, A., & Wendler, S. (2019). *AI adoption strategies*. Centre for Technology and Global Affairs: University of Oxford. Retrieved from https://www.ctga.ox.ac.uk/files/aiadoptionstrategies-march2019pdf

Donaldson, T., & Dunfee, T. (1994). Towards a unified conception of business ethics: Integrative Social Contracts Theory. *Academy of Management Review, 19*(2), 252–284.

Donaldson, T., & Dunfee, T. (1995). Integrative Social Contracts Theory: A communitarian conception of economic ethics. *Economics and Philosophy, 11*(1), 85–112.

Donaldson, T., & Dunfee, T. W. (1999). *Ties that bind: A social contracts approach to business ethics*. Cambridge: Harvard Business School Press.

Dunfee, T. W. (2006). A critical perspective of integrative social contracts theory: Recurring criticisms and next generation research topics. *Journal of Business Ethics, 68*(3), 303–328.

Estreder, Y., Rigotti, T., Tomás, I., & Ramos, J. (2020). Psychological contract and organizational justice: The role of normative contract. *Employee Relations: The International Journal, 42*(1), 17–34.

European Commission. (2020). *On artificial intelligence—A European approach to excellence and trust* [White paper]. Brussels. Retrieved from https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., et al. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines, 28*(4), 689–707.

Guenole, N., & Feinzig, S. (2018). *The business case for AI in HR: With insights and tips on getting started*. IBM Smarter Workforce Institute. Retrieved from https://public.dhe.ibm.com/common/ssi/ecm/81/en/81019981usen/81019981-usen-00_81019981USEN.pdf

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines, 30*, 99–120.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399.

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals, 14*(1), 366–410.

Laulié, L., & Tekleab, A. G. (2016). A multi-level theory of psychological contract fulfillment in teams. *Group & Organization Management, 41*(5), 658–698.

Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with machines: The impact of algorithmic, data-driven management on human workers. In *Proceedings of the 33rd Annual ACM SIGCHI Conference* (pp. 1603–1612). Seoul, South Korea: ACM Press.

Marr, B. (2017, October 16). *The 4 Ds of robotization: Dull, dirty, dangerous and dear*. Forbes. Retrieved from https://www.forbes.com/sites/bernardmarr/2017/10/16/the-4-ds-of-robotization-dull-dirty-dangerous-and-dear/#261e742b3e0d

Microsoft. (n.d.). *Responsible AI*. Retrieved from https://www.microsoft.com/en-us/ai/responsible-ai

Morrison, E. W., & Robinson, S. L. (1997). When employees feel betrayed: A model of how psychological contract violation develops. *Academy of Management Review, 22*, 226–256.

Norwegian Artificial Intelligence Research Consortium (NORA). (2020). *Norway's first National Strategy for Artificial Intelligence launched*. Retrieved from https://www.nora.ai/news-and-events/news/norway's-first-national-strategy-for-artificial-in.html

Obushenkova, E., Plester, B., & Haworth, N. (2018). Manager-employee psychological contracts: Enter the smartphone. *Employee Relations, 40*(2), 193–207.

Orlikowski, W. J., & Gash, D. C. (1994). Technological frames: Making sense of information technology in organizations. *ACM Transactions on Information Systems, 12*(2), 174–207.

Pal, A., & Ojha, A. (2017). Institutional isomorphism due to the influence of information systems and its strategic position. In *Proceedings of the 2017 ACM SIGMIS Conference on Computers and People Research* (pp. 147–154). https://doi.org/10.1145/3084381.3084395

Partnership on AI. (n.d.). *Tenets.* Retrieved from https://www.partnershiponai.org/tenets/

Parzefall, M., & Hakanen, J. (2010). Psychological contract and its motivational and health-enhancing properties. *Journal of Managerial Psychology, 25*(1), 4–21.

Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology, 20*(1), 5–14.

Robbins, S. (2019). AI and the path to envelopment: Knowledge as a first step towards the responsible regulation and use of AI-powered machines. *AI & Society*. https://doi.org/10.1007/s00146-019-00891-1

Rousseau, D. (1995). *Psychological contracts in organizations: Understanding written and unwritten agreements*. Thousand Oaks: Sage.

Rousseau, D., Hansen, S., & Tomprou, M. (2018). A dynamic phase model of psychological contract processes. *Journal of Organizational Behavior, 39*(9), 1081–1098.

Sainato, M. (2020, February 5). *'I'm not a robot': Amazon workers condemn unsafe, grueling conditions at warehouse*. The Guardian. Retrieved from https://www.theguardian.com/technology/2020/feb/05/amazon-workers-protest-unsafe-grueling-conditions-warehouse

Schalk, R., & Rousseau, D. M. (2008). Psychological contracts in employment. In Y. Altman, F. Bournois, & D. Boje (Eds.), *Sage library in business and management: Managerial psychology* (pp. 242–255). Sage.

Selbst, A., boyd, D., Friedler, S., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In: *ACM Conference on Fairness, Accountability, and Transparency* (FAT* 2018).

Sharkey, N. E. (2013). The evitability of autonomous robot warfare. *International Review of the Red Cross, 94*(886), 787–799.

Smith, N. C. (2000). Social marketing and social contracts: Applying integrative social contracts theory to ethical issues in social marketing. In A. Anderson (Ed.), *Ethics in social marketing*. Washington, DC: Georgetown University Press.

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review, 61*(4), 15–4228.

Telenor Group. (2019a). *Negotia and Telenor present Norway's first declaration for responsible use of AI to Digitalisation Minister Nikolai Astrup*. Retrieved from https://www.telenor.com/media/press-release/negotia-and-telenor-present-norways-first-declaration-for-responsible-use-of-ai-to-digitalisation-minister-nikolai-astrup

Telenor Group. (2019b). *Three ways we can shape the Workforce of the Future*. Retrieved from https://www.telenor.com/the-workforce-of-the-future/

Telenor Group. (2020). *Telenor sets a good example on AI upskilling*. Retrieved from https://www.telenor.com/telenor-sets-a-good-example-on-ai-upskilling/

Telenor Group. (n.d.). *Artificial intelligence*. Retrieved from https://www.telenor.com/innovation/artificial-intelligence/

Thompson, J. A., & Hart, D. W. (2006). Psychological contracts: A nano-level perspective on social contract theory. *Journal of Business Ethics, 68*(3), 229–241.

Treem, J. W., Dailey, S. L., Pierce, C. S., & Leonardi, P. M. (2015). Bringing technological frames to work: How previous experience with social media shapes the technology's meaning in an organization: Bringing technological frames to work. *Journal of Communication, 65*(2), 396–422.

Van Buren, H. J. (2000). The bindingness of social and psychological contracts: Toward a theory of social responsibility in downsizing. *Journal of Business Ethics, 25*(3), 205–219.

Walsh, T., Levy, N., Bell, G., Elliott, A., Maclaurin, J., Mareels, I., & Wood, F. (2019). *The Effective and ethical development of Artificial Intelligence* (p. 250). ACOLA. Retrieved from https://acola.org/wp-content/uploads/2019/07/hs4_artificial-intelligence-report.pdf

Wellman, M. P., & Rajan, U. (2017). Ethical issues for autonomous trading agents. *Minds and Machines, 27*(4), 609–624.

Winfield, A., Katina, M., Pitt, J., & Evers, V. (2019). Machine ethics: The design and governance of ethical AI and autonomous systems. *Proceedings of the IEEE, 107*(3), 509–517.

World Economic Forum. (2018). *The future of jobs report 2018*. Geneva: Centre for the New Economy & Society. Retrieved from http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf

Wright, S. A., & Schultz, A. E. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. *Business Horizons, 61*(6), 823–832.

Yaxley, M. (2019, November 9–10). Training key to success in a tech-intense world. In *The Australian* (p. 48).

Zhao, H., Wayne, S., Glibkowski, B. C., & Jesus, B. (2007). The impact of PC breach on work-related outcomes: A meta-analysis. *Personnel Psychology, 60*(3), 647–680.