Journal of Business Ethics
https://doi.org/10.1007/s10551-018-3887-1

ORIGINAL PAPER

Corporate Agency and Possible Futures

Tim Mulgan[1,2]

Received: 13 January 2017 / Accepted: 21 April 2018

Abstract
We need an account of corporate agency that is *temporally robust*—one that will help future people to cope with challenges posed by corporate groups in a range of credible futures. In particular, we need to bequeath moral resources that enable future people to avoid futures dominated by corporate groups that have no regard for human beings. This paper asks how future philosophers living in broken or digital futures might re-imagine contemporary debates about corporate agency. It argues that the only temporally robust account is *moralised extreme collectivism*, where full moral personhood is accorded (only) to those corporate groups that are reliably disposed to respond appropriately to moral reasons.

Keywords Collective agency · Corporate agency · Corporate responsibility · Future people · Climate change · Broken world · Virtual reality · Artificial intelligence

In this paper, I ask how future people might rethink the moral status of corporate groups, and what we can learn from reflection on their moral thinking. "Why Should We Think About Possible Futures?" section explains why we should care about possible futures, "The Present Debate About Corporate Agency" section introduces the current debate about corporate agency, while "Broken Futures" and "Digital Futures" sections argue that people living in broken and digital futures will think very differently about corporate agency.

Why Should We Think About Possible Futures?

There is a substantial philosophical literature on the moral status of groups and collectives (e.g. French 1979; Velasquez 2003; Copp 2007; List and Pettit 2011; Hess 2013; Hindriks 2014; and the other works cited below). I focus here on groups that have both a structure and a decision-making process, as opposed to merely aggregate groups such as 'the people currently in Heathrow Terminal 5'. I refer to these as 'corporate groups' (Tollefsen 2015, p. 3).[1] Our social world includes many corporate groups, such as business enterprises, legally incorporated entities, bureaucratic organisations, public bodies, clubs, societies, juries, judicial panels, or committees. We acknowledge the existence of corporate groups; we enter into contractual relationships with them; we recognise their legal personality; we accord them certain rights; we attribute goals, desires, and intentions to them; and we hold them responsible for their actions or inactions.

Why should we think about possible futures? In particular, why should we care what future people might think about corporate groups? I argue in "Broken Futures" section that people in possible *broken* futures will be more likely (for philosophical reasons) to accept the reality of corporate moral agency, but also more reluctant (for historical, philosophical, and social reasons) to recognise the moral agency of corporate groups that cannot prove themselves to be morally reliable. In "Digital Futures" section, I argue that people in possible *digital* futures will acknowledge uploaded or artificial digital beings as *moral*

✉ Tim Mulgan
tpm6@st-andrews.ac.uk

[1] Department of Philosophy, University of St Andrews, Edgecliffe, The Scores, St Andrews KY169AL, UK

[2] Department of Philosophy, University of Auckland, Private Bag 92019, Auckland, New Zealand

[1] While "corporate group" is slightly clumsy, alternatives such as "collective" or "corporate agent" bias the debate in favour of a collectivist interpretation. Note that "corporate group" is broader than "corporation", which refers only to legally incorporated entities. I use "business enterprise" to cover all firms whether legally incorporated or not. (I am grateful to two anonymous reviewers for pressing me to clarify my terms here.)

Published online: 03 May 2018

persons, and that this recognition will lead them both to acknowledge some corporate groups as persons, but also to be even more wary of morally unreliable corporate groups. Suppose you accept all these speculations about future people's beliefs. Why should you care? What impact should future beliefs have on our *current thinking* about corporate groups?

Your initial answer may depend on your methodological priorities. Philosophical accounts of corporate groups serve many different purposes. I distinguish three alternative approaches: *realists* seek to *explore* a metaphysical *reality* that exists independently of our aims and practices; *pragmatists* ask which account of corporate groups best serves our *practical purposes*; and *interpretivists* interpret our *current social practice* of holding corporate groups to account.[2]

Pragmatists will definitely find my project important. If we seek to develop and bequeath an account of corporate groups that serves our practical purposes, and if we care about future people at all, then we need to know what they will find useful or credible. This is especially true for consequentialists, who seek to maximise well-being into the distant future.

Knowledge of possible futures is also relevant for interpretivists. Our practice of dealing with corporate groups is part of our general practice of holding one another to account. And the latter contains several future-directed elements. Our moral norms include many explicitly future-directed obligations: to future people in general, to our descendants, to future inhabitants of this place, to future citizens of this polity, to future stakeholders in this corporate group, and so on. And any social practice implicitly regards itself as extending indefinitely into the future. The best *interpretation* of our present practice will be sensitive to future threats.

Both pragmatists and interpretivists thus seek a *temporally robust* account of corporate groups: one that provides the inhabitants of a broad range of credible futures with the conceptual resources they need to recognise, constrain, and hopefully avoid the particular threats that are most salient to them. At the very least, we must not bequeath a philosophical story about corporate groups that leaves future people unable to recognise the most worrying threats or prevents them from addressing them.

Realists, by contrast, may regard speculation about possible futures as irrelevant.[3] What matters is the *truth* about corporate groups, not what anyone believes. Discovering future beliefs should not affect present beliefs, because we have no reason to regard future people as more reliable than ourselves. And, of course, we are really dealing here with present *predictions* about future beliefs, not established facts about those beliefs. Even if future people *will* know better, we cannot non-question-beggingly help ourselves to future metaphysical discoveries!

I reply that even realists *should* care about possible futures, for two reasons. First, possible futures are relevant to realists *qua realists* because they transform philosophical thought experiments into real cases. In the rest of this paper, I discuss societies where self-sufficient survival is impossible, Rawlsian favourable conditions no longer apply, scarcity is ubiquitous, and historical justifications of property rights are no longer credible; and futures containing AIs capable of passing the Turing Test, outwitting humans, or forming corporate groups as complex as human brains. All these presently imaginary scenarios are more significant when we realise they might actually happen. This is because, as most realists will agree, even if we *ideally* seek to accommodate all *conceivable* cases, any *acceptable* metaphysical theory *must* accommodate all *actual* cases (past, present, and future). Any realist who thinks actual cases have *some* priority over imaginary ones should therefore pay *some* attention to possible futures.

More importantly, realists should care about possible futures because no realist is *only* a realist. Virtually no one believes that pragmatic and interpretive questions are unimportant, or that it doesn't matter how we actually treat corporate groups. Every realist philosopher is also a citizen in a society threatened by morally unreliable corporate groups, and a participant in our practice of holding corporate groups to account. Even if she had no interest in possible futures *qua realist philosopher*, she should still consider them in her other social roles.

Some realists will remain unsatisfied. If we foresee that future ethics will be (e.g.) racist, shouldn't we resist it rather than embracing it? I reply that a temporally robust ethic must satisfy three distinct criteria. (1) It must strike *future people* as plausible. (2) It must enable *them* to recognise and address salient ethical threats. (3) But it must also *not* strike *us—after reflection on possible futures*—as ethically unacceptable. If people in possible future F can only survive by embracing racism, tribalism, or xenophobia, then there is no ethic that is temporally robust with respect to F, and we should focus instead on trying to avoid F altogether. My

[2] I am indebted here especially to Tollefsen (2015). My tripartite distinction between realism, pragmatism, and interpretivism oversimplifies the vast literature on philosophical method. In particular, my 'pragmatism' is broader than the specific tradition of James or Dewey, and my 'interpretivism', while inspired by Strawson's work on "reactive attitudes" and Dennett's "intentional stance" (Tollefsen 2015, p. 119 and p. 97, respectively), is not necessarily committed to the details of their views.

[3] I am grateful to two anonymous reviewers for pressing me on this point.

task in "Broken Futures" and "Digital Futures" sections is to persuade the reader that the accounts of broken and digital future ethics developed in those sections *do* satisfy this third criterion.

I conclude that the search for a temporally robust account of corporate groups should interest everyone. To determine whether currently popular accounts *are* temporally robust, I explore ways that credible futures might transform current debates. Future people may see things differently because they recognise new moral individuals (such as digital people); accept different accounts of people's rights and responsibilities; regard different features of persons as important; favour different interpretations of contested concepts such as intentionality or consciousness; have different general philosophical priorities; or face different threats from corporate groups.

Any precise prediction about the future is almost certainly false. The future of human beliefs and behaviour is especially difficult to predict. I do not claim to provide a description of what will happen, or even of what is likely to happen. But I do present two credible futures whose challenges and threats are very real. If one account of corporate groups is much better equipped than its rivals to meet those challenges and threats, then it has a significant competitive advantage.

Some possible futures are more significant than others. If we are risk-averse, or if we give priority to the interests of the worst off, then we should pay disproportionate attention to futures whose inhabitants are worse off than ourselves, even if more optimistic futures are equally likely. We should also pay disproportionate attention to scenarios where our influence on future people's moral beliefs might have the greatest impact and where corporate groups themselves are most threatening.

I examine two possible futures: broken and digital. Each represents a wider class of possibilities characterised by either extreme scarcity or the emergence of non-moral corporate agents. I ask how people living in those futures might re-imagine our debates about corporate groups, and what we might learn from their future debates.

The Present Debate About Corporate Agency

Before turning to our two possible futures, I summarise the present philosophical debate about corporate groups, focusing on features that might change in the future.

Individualism Versus Collectivism

The debate about corporate agency is structured by a tension between two appealing intuitions (Hindriks 2014, p. 1566). (1) Some corporate groups *are* moral actors who should be held responsible for their actions—especially when no individual(s) can feasibly be held to account. (2) No corporate group is a natural moral person, and therefore no corporate group deserves all the moral rights enjoyed by human persons. As there is no agreed terminology, I stipulate a contrast between *corporate agency* and *corporate personhood* to capture this distinction.

Talk of corporate "agency" and "responsibility" is ambiguous. It combines ontological claims about the intentional and causal capacities of corporate groups with normative claims about their moral status. The ontological claim is that some corporate groups are *independent* agents whose beliefs, intentions, and actions cannot be entirely reduced to those of their individual members. Independent agents are *causally* responsible for their actions. (The classic example is where a committee does X because it believes Y, even though none of its members either do X or believe Y. See e.g. List and Pettit 2011, Chap. 6.) The normative claim is that these corporate groups are also *moral* agents who are *morally* responsible for their actions, can be expected to obey moral norms, enjoy some rights, and can therefore be wronged.

We can now distinguish four possible positions.[4]

1. *Individualism*. Corporate groups are not agents of any kind. The only agents are individuals human beings (and perhaps some other animals). All talk of corporate agency or personhood is metaphorical. Although we hold corporate groups "responsible" for legal or compensatory purposes, corporate groups themselves cannot bear genuine *moral* responsibility. It makes no sense to literally blame a corporate group.
2. *Minimal collectivism*. Some corporate groups are independent agents who act in the world, but none are moral agents or moral persons.
3. *Moderate collectivism*. Some corporate groups are moral agents, but none are fully fledged moral persons. Corporate groups enjoy some rights (e.g. property, contract) and can be held responsible for their actions. But they do not enjoy human rights.
4. *Extreme collectivism*. Some corporate groups are both moral agents and moral persons. They have the same moral status as human beings and enjoy analogous rights.

Individualists reject our first intuition; extreme collectivists reject the second; while minimal and moderate collectivists seek to reconcile the two intuitions. Collectivism allows us to hold corporate groups responsible even when no individual is responsible. Individualists, by contrast, must

[4] I borrow the individualism/collectivism distinction from Velasquez (2003), p. 38. The minimal/moderate/extreme distinction is my own.

accept that, if no individual is responsible, then no one is responsible (Velasquez 2003 explicitly bites this bullet).

I can find no mainstream theorist who defends extreme collectivism.[5] Even those who defend corporate "personhood" explicitly distance themselves from full moral personhood in my sense. (e.g. French 1979; Goodpaster and Matthews 1982; List and Pettit 2011; Hindriks 2014, p. 1567; Hess 2013, p. 319.) No one argues, for instance, that corporate groups should enjoy the right to vote, the right to life, or the right not to be owned by other persons[6] (Hindriks 2014, p. 1576 refers to these rights as "the problematic trio"). Minimal collectivism is also not a prominent view. The primary motivation for treating corporate groups as *independent* agents is the desire to hold them *morally* responsible for their actions. By contrast, moderate collectivism has many proponents (e.g. French 1979; Copp 2007; List and Pettit 2011; Hess 2013, p. 14; Pettit 2014), as does individualism (e.g. Narveson 2002; Velasquez 2003; Ashman and Winstanley 2007; Bevan and Corvellec 2007; Miller and Makela 2005). In this paper, my primary interest is in the moral status of corporate groups. I therefore focus on the debate between individualism and moderate collectivism—although extreme collectivism re-emerges in "Digital Futures" section.

Individualists argue that moderate collectivism is an unstable position (e.g. Velasquez 2003; Rovane 2014). Once we recognise corporate groups as moral agents, we are on a slippery slope to recognising their full moral personhood. But it would be absurd to treat corporate groups as persons. Therefore, we must reject moderate collectivism. Moderate collectivists deny that moral agency inevitably leads to moral personhood, while extreme collectivists embrace this "slippery slope".

Defending Moderate Collectivism

Moderate collectivists must defend the *moral agency* of corporate groups while rejecting full corporate *moral personhood*. They typically proceed as follows:

1. Identify paradigmatic moral persons;
2. List similarities and dissimilarities between paradigmatic moral persons and corporate groups;
3. Argue that the similarities show that some corporate groups *are* moral agents, who may enjoy property rights and contractual rights;
4. Argue that the dissimilarities show that all corporate groups are *not* moral persons and therefore don't enjoy distinctively *human rights* such as life, liberty, and political participation.

All sides agree that our paradigm moral persons are normally functioning adult human beings who are both moral agents and moral persons. They cause things to happen, have obligations, can be held responsible for their actions, and enjoy both agency rights (e.g. they can own property and enter into contracts) and distinctively human rights (e.g. they should be entitled to vote, they cannot be enslaved or unlawfully killed, etc.).

Debate persists because we cannot agree either on the essential distinctive features of paradigmatic moral persons, nor on whether any corporate group shares those features. Agency and personhood both involve a web of inter-related contested concepts. The following features of human adults have been identified as either necessary or sufficient for either moral agency or personhood: having a soul (cf. Pettit 2014, p. 1642, attributing the view to Pope Innocent IV); being embodied (Hess 2013); being a biological organism (Velasquez 2003; Miller and Makela 2005); being sentient, aware, self-aware, conscious, or self-conscious (Ashman and Winstanley 2007; Bevan and Corvellec 2007; Hussain and Moriarty 2016); being free, autonomous, or self-directed (Copp 2007; List and Pettit 2011; Hess 2014); possessing intentionality (Ashman and Winstanley 2007; Hindriks 2014, p. 574; Velasquez 2003; French 1979; List and Pettit 2011); being rational or responsive to reasons (French 1979; Bratman 2000; Pettit 2007, List and Pettit 2011; Hess 2014); being moral, responsive to moral reasons, empathetic, or other-regarding (Bevan and Corvellec 2007; Velasquez 2003; List and Pettit 2011).

Moderate collectivists offer stricter conditions for personhood than for moral agency. For instance, perhaps personhood demands sentience or consciousness, while intentionality or rationality is sufficient for moral agency (e.g. French 1984; List and Pettit 2011; Hindriks 2014, p. 1567; List 2016). By contrast, individualists often elide agency and personhood, defending similar conditions for both (e.g. Velasquez 2003; Ashman and Winstanley 2007; Bevan and Corvellec 2007).

Even philosophers who agree on the criteria for agency or personhood often disagree on their application to corporate groups. Some properties clearly rule out corporate groups. Such groups are not embodied biological organisms, nor do

[5] One possible exception is French (1984), although French himself adopts a weaker interpretation in his own more recent work. (I am grateful to an anonymous reviewer for alerting me to this.)

[6] Hasnas (2013) *presents* an argument that corporations should enjoy voting rights. But this argument is explicitly presented as a reductio ad absurdum of its premises. (I owe this reference to an anonymous reviewer.) Similarly, Kusch argues that it is inconsistent for List and Pettit to allow corporations to be enslaved! (Kusch 2014, pp. 1597–1598).

they have souls.[7] If these things are essential to moral agency or personhood, then corporate groups cannot qualify. Collectivists must deny that these features are necessary—arguing either that they are incidental to moral agency (e.g. organic embodiment), or that human do not possess them and therefore they cannot be the basis of *our* agency (e.g. souls).

All the other items on our list are sites of intense philosophical controversy. Collectivists typically focus on intentionality, rationality, and responsiveness to reasons. They defend particular *interpretations* of these concepts that allow for both human and corporate moral agents, while individualists favour interpretations that exclude non-human agents. In particular, collectivists favour *functionalist* views where "mental states are to be defined in terms of what they do rather than in terms of their physical make-up" (Tollefsen 2015, p. 69. See also, e.g.; French 1979; List and Pettit 2011; Pettit 2014). Agency is an organisational property, not a matter of having the correct biological substrate. And "if a biological or vital organism can be functionally organized so as to meet the constraints of agency, why can't an artificial entity be organized in that way as well? Why not a suitably engineered robot, for example? And why not a suitably organized group of individuals?" (Pettit 2014, p. 1645) [We return to robots in "Digital Futures!" section.] By contrast, individualists favour *phenomenological* or *organic* accounts where intentionality demands either conscious awareness or specifically human embodiment (e.g. Ashman and Winstanley 2007 drawing on Husserl; Bevan and Corvellec 2007 on Levinas; Velasquez 2003 on the Thomist tradition; and Miller and Makela 2005 on "common-sense").

Both sides typically assume that corporate groups are *not* conscious. They have no experiences of their own, no inner goals, no qualia. There is nothing it is like to be a corporate group.[8] Individualists, who regard consciousness as essential to *any* form of agency, conclude that corporate groups cannot even be independent agents (let alone *moral* agents). Moderate collectivists reply that, while lack of consciousness is no barrier to moral *agency*, it does explain why corporate groups are not *persons*, and therefore cannot enjoy distinctively human rights. We can attribute responsibility, property, and contract to any moral agent. But life, political participation, and freedom from slavery only matter to conscious beings.[9]

Hindriks notes a tension within moderate collectivism at this point. If one embraces a functionalist analysis of intentionality, why not endorse an analogous account of consciousness? But then: "If sentience is a functional property, it cannot be tied uniquely to a particular kind of matter". (Hindriks 2014, p. 1582) I argue in "Digital Futures" section that, for similar reasons, future people may find it much harder to deny that corporate groups are conscious.

Responding to Moral Reasons

The final criterion on our list is responsiveness to *moral* reasons. There is broad agreement that, in order to act in the world, every moral agent must respond to *some* reasons: an individual or corporate group whose behaviour was random, mechanical, or purely instinctual would not be any kind of *agent*. The interesting controversy concerns *moral* reasons.

To explore the significance of moral reasons, we must first introduce two further distinctions. (Cf. List and Pettit 2011, p. 158.) One is between *recognition* of moral reasons and *appropriate responses* to those reasons. The other (orthogonal) distinction is between *ability* and *reliability*. A putative agent might be *able* to recognise moral reasons, or to respond appropriately, without being reliably disposed to *exercise* those abilities.

We can now separate several different *moral* criteria for moral agency or personhood. The least demanding moral criterion would only insist on the ability to recognise moral reasons. Next, we could insist on the *ability* to respond appropriately. (e.g. List and Pettit 2011, Chap. 7) The most demanding moral criterion would require *reliably appropriate response*.

I will concentrate on the strongest reliability-based criterion, partly because it is both the most interesting and the most controversial, but also because it is most likely to come to the fore in possible futures. I will refer to *collectivists* who endorse that condition as *moralised collectivists*. (Moralised collectivism can be either extreme or moderate, depending on whether morally reliable corporate groups qualify as persons or only as moral agents.)

[7] Corporate groups made up of human beings are *constituted* by biological organisms. So they are *embodied* in that sense. But no corporate group has the *organic unity* of a single organism. (I am grateful to an anonymous reviewer for pressing me on this point.)

[8] Collectivists who deny the possibility of group consciousness include Tuomela (2013), p. 52; Theiner (2014); Tollefsen (2015), p. 63; List (2016). Two rare exceptions who do not rule out this possibility are Huebner (2014), p. 120, and Schwitzgebel (2015). Note that what is at issue here is *phenomenal* consciousness. Many collectivists argue that corporate groups can enjoy weaker non-experiential alternatives such as "access consciousness" or "consciousness as awareness". (Tollefsen 2015, p. 52; List 2016, p. 6).

[9] The growing literature on political CSR (corporate social responsibility) may seem to be an exception, because it argues that corporations (in particular) *should* be involved in politics (e.g. Scherer and Palazzo 2011). However, that literature addresses the distinct question of how corporations should contribute to political institutions designed to further the interests *of individuals*, rather than asking whether corporations possess a *moral* right to participate in political decision-making to further *their own* interests. (Here I follow the critique in Hussain and Moriarty (2016). I owe this reference to an anonymous reviewer.)

Moral conditions are often introduced in response to the problem of psychopathic (or otherwise morally unreliable) corporate groups. Normally functioning human adults recognise, and respond to, moral reasons. They accept that the interests of others constrain their pursuit of their own goals. But some human adults cannot recognise moral reasons or other-regarding constraints. At the risk of over-simplification, let us call such people "psychopaths". (On the broader philosophical significance of psychopathy, see e.g. Malatesti 2010.) A familiar theme in the recent literature is that many actual corporate groups—and especially for-profit business enterprises—are closer to psychopaths than to normally functioning human adults (e.g. Bakan 2004, p. 56ff). They can act, intend, plan, and display instrumental rationality. But they feel no empathy, cannot recognise the moral importance of others' points of view, and cannot recognise (and therefore cannot respond appropriately to) moral reasons. Indeed, psychopathic behaviour is encouraged by modern theories of the firm that insist business enterprises should single-mindedly maximise returns to shareholders! (Nor is legal incorporation any barrier to psychopathy, because modern corporations can pursue any legal purpose.)

No one denies that human psychopaths are independent agents who act in the world. Moralised collectivism is thus not a viable option for the *minimal* collectivist. If non-psychopathic corporate groups are independent agents, then so are psychopathic ones. But if the ability to recognise moral reasons is an essential feature of *moral* agency, then psychopathic corporate groups are not moral agents. The most minimal moral condition would rule out psychopaths. But we need stronger reliability-based criteria to address the threat posed by *all* corporate groups that (for whatever reason) are not reliably disposed to respond appropriately to moral reasons. Psychopathic corporate groups are merely one striking subset.

Individualists will reply that the best defence against any abuse of corporate power is to insist that all talk of "corporate action" is merely a convenient short-hand, that no corporate group has *any* agency over-and-above the agency of its members, that corporate groups don't *really* have any rights, and that individuals should *never* be allowed to evade responsibility by shifting it onto corporate groups. Moralised collectivists will object that this solution is not temporally robust. As corporate groups become more sophisticated and ubiquitous, it will become increasing impossible to insist that *no* corporate group is a moral agent. Our only hope is to draw a principled distinction between morally reliable corporate groups and unreliable ones, and then prevent unscrupulous individuals from shifting responsibility onto the latter *by denying that morally unreliable agents are moral agents at all*.

Moralised collectivism is a minority position in the present debate. But it is endorsed or suggested by some moderate collectivists. (Cf. List and Pettit 2011, p. 159; Hess 2013; Tollefsen 2015, Chap. 6.) I shall argue that, in the future, this minority position will become much more prominent. Indeed, I will argue that moralised collectivism is the only temporally robust position.

The Interpersonal Dimension

These differences between individualists and collectivists are relatively clear-cut. The next point of difference is harder to pin down in particular cases. But it seems to me to play a key role in the dialectic between the two camps. This new disagreement is not about a specific criterion. Rather, it affects the interpretation of several independent criteria.[10]

Collectivists place a much greater emphasis on the *interpersonal dimension* of agency. This is partly a matter of different starting points. Collectivists focus on how we recognise one another as agents. Any acceptable criterion of agency must be publically assessable, and therefore criteria based on inner mental states are prima facie inferior to those based on observable interpersonal behaviour. By contrast, individualists start from each individual's experience of her own moral personhood. They are therefore innately suspicious of public interpersonal criteria precisely *because* they could come apart from the agent's private inner states.

This difference in starting points tracks a deeper disagreement. Collectivists often favour conceptions of intentionality or rationality that make these contested concepts *intrinsically* interpersonal. The ability to interact successfully with others, and to make oneself understood by them, is an essential feature of either independent agency or moral agency. (The clearest example is Philip Pettit, whose accounts of corporate agency in Pettit 2007, 2014; List and Pettit 2011 build on the interpersonal account of the capacity for thought and concept possession set out in Pettit 1993.) Individualists' lists of essential features are typically more self-contained, as are their interpretations of contested concepts such as intentionality or rationality. A socially isolated human agent could still possess a soul, be an embodied animal, or be conscious.

For our purposes in this paper, these disputes about interpersonal relations matter for two reasons. First, the collectivist shift from unverifiable inner states such as consciousness to observable interpersonal interactions makes it easier for corporate groups to be publically recognised as moral agents

[10] This section explores general tendencies, not universal differences. I do not claim that every collectivist emphasises the interpersonal dimension of agency more than every individualist. I claim instead only that, in general, collectivists give this dimension greater weight than individualists. (I am grateful to an anonymous reviewer for pressing me on this point.)

especially if it is in their own interest to be so recognised.[11] Second, I argue below that the interpersonal dimension itself becomes more important in some significant possible futures. Because collectivism already emphasises that dimension, it thus enjoys a comparative advantage in those futures.

Three Philosophical Methods

Disagreement about the significance of interpersonal interactions shades into a broader disagreement about philosophical methodology. As I outlined in "Why Should We Think About Possible Futures?" section, philosophical accounts of corporate moral agency and personhood can be realist, pragmatist, or interpretivist.

For the realist, the philosopher's primary task is to determine the (metaphysical) facts (e.g. Velasquez 1983, 2003). Normative questions arise only *after* the metaphysical facts are settled, with the moral status of corporate groups determined by their metaphysical nature. If morally unreliable corporate groups meet the necessary conditions for moral agency, then we should acknowledge the *fact* that they are moral agents. Conversely, if no corporate group (not even a morally reliable one) satisfies those conditions, then we should acknowledge that *as a matter of fact* there are no corporate moral agents.

The pragmatist, by contrast, insists that both metaphysical and normative questions are settled soley by reference to *our interests* (e.g. Ashman and Winstanley 2007; Donaldson 1982; Dubbink and Smith 2011; Scherer and Palazzo 2011; Werhane 1985. For critique, see Hasnas 2013). We should acknowledge corporate groups as independent and/or moral agents, recognise their rights, hold them to account, and so on, if *and only if* doing these things best promotes what matters *to us*.

Some philosophers combine realism and pragmatism, often reserving the latter for questions about moral status. List and Pettit's moralised collectivism nicely illustrates this possibility (List and Pettit 2011, Chap. 7). While their account of independent agency can be interpreted as realist (or interpretivist), their moral condition is clearly *pragmatist*. They argue that we should restrict moral agency to corporate groups that are morally reliable, not because this tracks some underlying metaphysical reality, but rather because it best serves the *practical* purpose of advancing the interests of human individuals. List and Pettit adopt "a normatively individualist framework … that treats only individual people … as ultimate units of moral significance, while assigning only derivative moral significance to [corporate groups]". (List 2016, pp. 20–21) They also take a broadly consequentialist approach—seeking rules governing collective moral agency that will maximise human welfare into the future.

We now confront another potential ambiguity within collectivism. To *recognise* another's moral agency or personhood can mean *either* to acknowledge that the other *is* (independent of one's recognition) a moral agent or person, *or* to decide to treat the other as a bearer of certain moral rights and duties. Realists typically use 'recognise' in the first sense, while pragmatists have in mind the second sense. In what follows, to remain neutral between realism and pragmatism, I typically speak generically of *recognising* moral agency or personhood. But it should always be remembered that realists and pragmatists will interpret this recognition in quite different ways.

The interpretivist takes an intermediate position (e.g. Tollefsen 2015. For critique, see Huebner 2014). Like the realist, she seeks to describe not prescribe. But her descriptions are internal to our social practices. The interpretivist asks whether individualism or collectivism offers the better interpretation of our practice of ascribing moral agency and moral responsibility to both individuals and corporate groups. We should embrace whatever account of corporate groups best fits that practice. Realism and pragmatism both take seriously the possibility that our current practices could be radically deficient—either because they fail to reliably track metaphysical reality or because some feasible alternative would better serve our practical goals. By contrast, the interpretivist takes our current practice as given. There is no external standard, and an existing practice can only be judged to be deficient according to standards internal to itself.

Logically speaking, these methodological (realist vs. pragmatist vs. interpretivist) and substantive (individualist vs. collectivist) divisions are orthogonal. Realists could endorse either individualism or collectivism, as could pragmatists and interpretivists. And most individualists and collectivists offer a variety of realist, pragmatist, and interpretivist arguments for their respective substantive positions. However, there are some notable correlations between substance and method. Individualists are more likely to emphasise realist arguments, while collectivists are more likely to offer pragmatist or interpretivist ones. In particular, individualists often insist, on metaphysical grounds, that it simply makes no sense to say that corporate groups are agents *of any kind*; while moderate collectivists focus on the (negative) practical implications of denying that corporate groups are *moral agents*.

[11] While interpersonal interactions within a corporate group can be hidden from public view, they are at least observable in principle—unlike individual phenomenological states. Corporate groups therefore *can* lay their inner workings open to outside inspection if they have sufficient incentive to do so. Other agents can then ensure transparency by providing such incentives. (I am grateful to an anonymous reviewer for pressing me on this point.)

Present philosophers' views about corporate groups depend on their prior views about *human* moral agency and personhood, and on their philosophical priorities. Future people may see things differently because they recognise new paradigmatic moral people; accept different accounts of their rights and responsibilities; regard different features of persons as important; favour different interpretations of contested concepts such as intentionality or consciousness; have different general philosophical priorities; or face different threats from corporate groups. For instance, in a digital future where both adult humans *and digital being* are paradigm moral persons, credible criteria for moral agency or personhood must cover persons of both kinds.

We now turn, finally, to our exploration of broken and digital futures. Over the next two sections, I defend two general conclusions: reflection on future ethics offers us new resources to defend ourselves against morally unreliable corporate groups, and the only temporally robust position is moralised extreme collectivism. I also argue that the future of ethics is *path dependent*. How people think in the distant future will be influenced by the particular ethical tools they inherit from intermediate futures. In particular, digital futures look very different when approached from the standpoint of an already broken world.

Broken Futures

Imagine a future broken by climate change, where a chaotic climate makes life precarious, Rawlsian "favourable conditions" (Rawls 1971, p. 178) no longer apply (i.e. it is no longer possible to meet all basic needs and respect all basic liberties), and our affluent way of life is no longer an option. This is one *credible* future. No one can reasonably be confident it won't happen. It involves no outlandish claims, scientific impossibilities, or implausible expectations about human behaviour. Climate change—or some other disaster—might produce a broken future.

To make our discussion more concrete, I also assume that the broken future results primarily from anthropogenic climate change to which the wealthiest individuals, nations, *and corporate groups* have disproportionately contributed. The inclusion of corporate groups is not ad hoc. While discussion of differential responsibility for anthropogenic climate change typically focuses on the causal and moral responsibilities of wealthier individuals and/or nations, there is considerable evidence that the world's most powerful corporate groups are also casually implicated. For instance, Heede (2014) argues that 63% of all emissions of CO_2 and other greenhouse gases between 1751 and 2010 can be traced to 90 international entities (investor-owned companies, state-owned enterprises, or current and former centrally planned states), while the ten largest investor-owned companies alone contributed 15.8%. Heede himself concludes that this analysis "suggests a somewhat different, and perhaps useful way to consider responsibility for climate change" (Heede 2014, p. 235). In addition to their direct emissions, some corporate groups also indirectly exacerbate climate change by undermining legislative proposals to curb green house gas emissions—via lobbying, promoting misinformation and uncertainty, or funding candidates, individuals, and organisations who are "sceptical" about climate change (Arnold 2016; Oreskes and Conway 2010). In the United States, in particular, Arnold argues that "a major cause of this failure [to enact legislation to curb green house gas emissions] has been corporate political activity intended to defeat such legislative efforts" (Arnold 2016, p. 233).

The threat of a broken future forces us to take our obligations to future people more seriously. People living *in* such a future must also rethink *their* basic ethical commitments. In this section, I argue that the moral resources of the broken future are also essential components of *our* best response to the threat of morally unreliable corporate groups.

Ethics in Broken Futures

Drawing on my own earlier work (Mulgan 2011, 2014, 2015a, 2015b, 2016a, 2016b, 2017, 2018), I now briefly sketch five key ways that the ethical outlook of a broken future society might differ from our own.

Rethinking Rights in an Age of Scarcity

In a broken future, scarcity of material resources (especially water) and an unpredictable climate create periodic population bottlenecks where not everyone can survive. (This is what the loss of Rawlsian favourable conditions *means*.) When *nothing* (not even bare survival) can be guaranteed to everyone, rights must either be abandoned or radically reinvented. Social survival in a broken world may require restrictions on personal liberty on a scale that people have only previously accepted in times of war or other temporary crisis. Private land and individual labour might be requisitioned to grow food; the use of fossil fuels for private purposes might be severely curtailed; and individual lifestyle choices—especially reproductive decisions—might be much more tightly regulated and constrained than we would accept. Our affluent liberal ethics, designed for a world of enduring favourable conditions and emphasising individual *rights*, is thus particularly *ill-suited* to a broken world. This is why the broken world is so ethically unsettling.

Abandoning Historical Entitlements

My broken future is disproportionately caused by the wealthiest individuals, nations, and corporate groups. Even if they

inherited legitimate property rights, morally responsible *future* individuals, nations, and corporate groups would be overwhelmed by an obligation to assist people whose plight is a direct consequence of the actions of their forebears or even (for intergenerational nations or corporate groups) their *own* past actions. No one in a broken world who inherited property holdings could use them entirely as they wished.

From Natural Rights to Consequentialism

Most natural rights theorists concede that *present* property rights are only justified if they benefit (or at least do not harm) *future* people. (For instance, many libertarians enshrine this commitment in Lockean provisos.) In our affluent world, this future-directed constraint lies in the background, because philosophers routinely take it for granted that future people will be better-off than present people. In a broken future, by contrast, it will move centre-stage. This has both practical and theoretical consequences. Practically, future philosophers will deny that our property rights were ever legitimate, because our exercise of those rights has left our descendants worse off due to climate change. They will therefore deny that individuals or corporate groups in their broken world could possibly have inherited any property rights from our affluent world. Theoretically, future philosophers will also reject natural rights theories that present rights as absolute side-constraints. They will instead favour forward-looking consequentialist accounts where rights are justified by future benefits and constrained by changing circumstances. At the very least, future philosophers will be much more sympathetic to consequentialism, and more suspicious of natural rights, than contemporary philosophers.

The Urgency of Cooperation

In a broken world, collective survival demands social cooperation on an unprecedented scale. Broken world thinkers will attach much greater significance to the ability to recognise, respect, and safeguard the long-term collective interests of human beings. Nurturing and developing this ability will be the central task of moral education and public institutions. This positive emphasis also has a negative flip-side. In a world where *every* inefficiency results in unnecessary deaths or increased risks of social collapse, and where the rights to life, liberty, and reproductive freedom of *morally reliable* persons are severely curtailed, the rights of *unreliable* human beings may be restricted even further. In our affluent liberal societies, we can afford to be tolerant, within limits, of psychopaths, refuseniks, and other unreliable individuals. Broken world dwellers, who regard such people as threats to the very existence of society, will be much more wary. Broken world society will therefore be both (a) much less permissive regarding reproductive, childrearing or other choices that could result in the emergence of morally unreliable individuals; and also (b) less tolerant of such individuals once they emerge.

The Changing Role of Philosophy

For several reasons, broken world philosophers will be more likely than present philosophers to prioritise pragmatism over realism. All broken world inhabitants make sacrifices for the common good more readily than we do. (Evidence from earlier eras when people often lived in less abundant circumstances strongly suggests that greater self-sacrifice is *possible*. And a broken world *society*—one whose foundations are not xenophobic or otherwise ethically unacceptable—is impossible without it!) Given their own grim history, future philosophers will also take their own intergenerational obligations much more seriously than we do, and place greater importance on collective and intergenerational projects. Philosophers are people too. In a broken world, where the stakes are always higher, philosophers will feel more pressure (from both internal and external sanctions) to consider the long-term real-world *consequences* of competing philosophical positions. This change in priorities reinforces the other distinctive features of broken world ethics—especially the focus on cooperation rather than self-reliance, the downgrading of individual rights, and the upgrading of forward-looking consequentialism.

Corporate Groups in Broken Futures

Two factors will influence broken world thinking about corporate groups. On the one hand, corporate groups will loom even larger in the broken world than in our affluent present. Because it demands unprecedented sacrifices for the common good, social survival in a broken world needs a *stronger* sense of collective solidarity than anything that exists in any contemporary Western society. Any broken future *society* will give a prominent role to corporate groups that fairly distribute scarce resources, allocate responsibilities, represent the interests of future people, and enforce entitlements. On the other hand, broken world philosophers will also be more concerned than us about the long-term threat posed by *morally unreliable* corporate groups—especially as they will be aware of the damage done to their world by the psychopathic business enterprises of our own age!

Being more inclined to pragmatism than present philosophers, broken world philosophers will be especially anxious to reduce the threat posed by morally unreliable corporate groups, without abandoning corporate groups altogether. How might they approach this task?

One option is to embrace *individualism*, and deny that any corporate group is an agent of any kind. Unfortunately, all the distinctive features of broken world philosophy count

very strongly against individualism. I argued in "Ethics in Broken Futures" section that broken world philosophy emphasises interpersonal cooperation and moral reliability, replaces backward-looking natural rights theory with forward-looking consequentialism, down-plays the significance of individual entitlements and self-reliance, and is more pragmatist than contemporary philosophy. This combination of factors strongly suggests that broken world *criteria* for agency and personhood will emphasise interpersonal interaction, publically assessable behaviour, and long-term consequences rather than the individual's private mental states and/or intrinsic biological or metaphysical properties.

This change in emphasis has two salient implications. Most obviously, we can expect a downgrading of precisely those criteria that corporate groups cannot *possibly* meet, such as ensoulment or embodiment. While some broken world philosophers will no doubt continue to regard consciousness, embodiment, or even ensoulment as important features of human agents, the increased emphasis on other criteria will diminish the *comparative* importance of these individualist-friendly features—both for individual philosophers and within the philosophical community as a whole.

More subtly, when they turn to contested concepts such as intentionality or rationality, broken world philosophers will favour *interpretations* that emphasise public behaviour and interpersonal interaction over private states or biological factors. In particular, they will prefer functionalist interpretations where "mental states are to be defined in terms of what they do rather than in terms of their physical make-up" (Tollefsen 2015, p. 69). As I argued in "Defending Moderate Collectivism" section, these are precisely the interpretations of contested concepts that are most favourable to corporate groups.

These theoretical considerations will make individualism less appealing to *all* broken future philosophers. Realists will reject individualism because it doesn't meet the criteria they regard as essential for metaphysical truth; interpretivists because it doesn't respect broken world practices that employ those criteria; and pragmatists because its lack of credibility means that individualism cannot provide a useful defence against the threat of morally unreliable corporate groups.

Broken world philosophers will favour collectivism over individualism. They will also favour *moralised* collectivism—recognising as moral agents only those corporate groups that *reliably respond appropriately to moral reasons*. There are several reasons for this. I argued in "Ethics in Broken Futures" section that broken world ethics is less tolerant of morally unreliable individuals than our contemporary affluent ethics. Morally unreliable *corporate groups* will be tolerated even less. No stable society could possibly persist unless normal patterns of moral development ensure that most human adults *do* recognise moral reasons and reliably respond appropriately. Any functioning moral code includes a presumption of moral reliability that treats human adults as innocent until they prove themselves guilty. (And proven psychopathic or morally unreliable *humans* cannot simply be ignored or destroyed without considerable moral cost—even in the harsh world of the broken future.) By contrast, if a corporate group's constitution allows it to pursue any legal purpose, while its management structure lacks any mechanism for recognising *and then responding appropriately to* moral reasons, then it is irredeemably psychopathic. (And a psychopathic *corporate group*, unlike its human equivalent, has no moral standing *at all*.) In a fragile world of extreme scarcity, anyone wishing to create a new corporate group—and have it recognised by others—must *first* demonstrate that it *will* be morally reliable.

Broken future pragmatists have an additional reason to favour moralised collectivism. In a world where personal liberty and privacy are generally more circumscribed, it will be much harder (thanks to both public institutions and social mores) to create a new corporate group *without* public scrutiny of its moral reliability. Like the creation of a new human being, creating a new corporate group will not be seen as a self-regarding private act! Moralised collectivism thus promises *better* protection against the emergence of morally unreliable corporate groups in a broken future than it does in our unbroken present. While individualism and non-moralised collectivism fare *worse* in the broken future, moralised collectivism fares better, especially by the pragmatist standards that future philosophers will privilege.

I conclude that, relative to our current debate, broken world philosophers will be more critical of both individualism and non-moralised collectivism, and more favourable towards moralised collectivism. We can expect that moralised collectivism to remain *moderate*. The transition to a broken world will not undermine the natural human reluctance to grant distinctively *human* rights to corporate groups. Indeed, the known consequences of granting such rights to some corporate groups in the past will be a stark warning! Extreme collectivism is likely to remain a minority position. However, as we shall see in "Digital Futures" section, the broken world could lead to other possible futures where even moderate collectivism becomes untenable and *extreme* collectivism is the only credible option.

Embracing Broken Future Ethics

I have argued that future philosophers will favour moralised collectivism over both individualism and non-moralised collectivism. If we seek a temporally robust account of corporate groups, this gives *us now* a reason to look more favourably on moralised collectivism. This reason is obviously most relevant to philosophers who endorse both consequentialism and pragmatism, because they already seek the account of

corporate groups that promises the best consequences into the distant future. But, as I argued in "Why Should We Think About Possible Futures?" section, anyone who cares about the future has some reason to embrace the ethical outlook of the broken future.

I would also argue that, whatever their starting-point, reflection on my broken future should make all philosophers more sympathetic to both pragmatism and consequentialism. My broken future reminds us that our present actions may have a very significant—and very negative—impact on future people who are worse off than ourselves. And the possible negative future impact of *present-day* corporate groups reminds us that ideas about corporate agency and responsibility are not harmless objects of philosophical speculation. Indeed, if we think of the resources of our world in intergenerational terms, then perhaps our world is already broken—the resources of the earth may already be too damaged to meet the basic needs of everyone who will live in the future. Perhaps the pragmatist consequentialist ethic of the broken future should become our ethic too.

Digital Futures

Imagine a *digital future* where flesh-and-blood humans have been replaced by digital beings—intelligent machines and/or digital copies of human brains (e.g. Blackford and Broderick 2014; Bostrom 2014; Chalmers 2010; Hauskeller 2013, p. 115–132; Mulgan 2014, 2016b, 2018). This is another credible future. No one can reasonably be confident it won't happen. We should be wary of breathless predictions of the imminent rise of super-intelligent machines (see, e.g. the critique presented in Floridi 2014). But confident pronouncements that artificial intelligence and digital uploading will forever remain engineering impossibilities are equally suspect. Computers continually confound their critics by performing tasks long deemed "impossible" ("No computer will ever play Checkers, or Chess, or Go; drive a car; recognise a face", etc.)

Digital futures could be especially appealing to people whose "real-world" alternative is already broken. A broken world might possess sufficient resources to upload, store, and "run" billions of minds, but not to maintain a comparable number of human beings. At the extreme, perhaps only digital beings can survive some catastrophe that will be fatal for all biological humans. The digital future would then be the only possible inhabited future.

The digital future introduces the threat of an *inhuman corporate future*, where there is no human participation in the dominant corporate groups at all. Every worker, manager, contractor, customer, board member, voter, regulator, office-holder, or share-holder is a digital being. If digital futures are credible, then so too are inhuman corporate ones. As their own numbers and processing speeds increase, digital beings can create new corporate groups at an ever-increasing rate.

The threat of morally unreliable corporate groups thus arises in the digital future in a very acute form. Any temporally robust account of corporate groups must cope with such futures. Unfortunately, the digital future is also especially destabilising for our present thinking about corporate groups—largely because it introduces *new paradigm moral persons*, namely digital persons.

Philosophical debate about digital beings mirrors the debate about corporate groups. We have four analogous substantive positions: exclusivism (only biological humans can be agents or persons), minimal inclusivism (digital beings can be independent agents but not moral agents or persons), moderate inclusivism (digital beings can be moral agents but not persons), and extreme inclusivism (digital beings can be full moral persons with human rights and morally significant interests). And we have the same three philosophical methodologies: realist, pragmatist, and interpretivist. Are we trying to discover the metaphysical truth about digital beings, decide how best to deal with them, or interpret an existing practice?

At present, digital beings are imaginary. Philosophical discussion is therefore predominantly realist. Most pragmatists are not (yet) worried about digital beings, and there is no extant 'practice regarding digital beings' to interpret. By contrast, in a future where digital beings are common-place, pragmatism and interpretivism will come to the fore. And they will both favour extreme inclusivism. Pragmatically, it will make little sense to deny the personhood of beings who are indistinguishable for all practical purposes from flesh-and-blood humans. And we can imagine a digital interpretivist paraphrasing Tollefsen's defence of collectivism about corporate groups: "I think it is clear that our lives would be greatly impoverished by relinquishing the reactive attitudes toward [digital beings]. … Eliminating our emotional responses to [digital beings] would eliminate the possibility of relationships with [digital beings] and relationships of this sort are a substantial part of society. Indeed, many of our relationships with [human] individuals seem to be dependent upon … relationships with [digital beings]". (Tollefsen 2003, pp. 229–230).

A future where some digital beings are recognised as full moral persons alongside adult human beings is definitely *epistemically* possible. No one can be confident it will not happen. This future is also especially troubling. Any temporally robust account of corporate groups must cope with it. I therefore stipulate that, in *my* digital future, digital beings *are* recognised as moral persons.[12]

[12] By insisting that a future where digital beings successfully match the cognitive achievements of human beings is even *possible*, don't I simply beg the question against philosophers such as John Searle who

In this section, I first explore the impact within digital future ethics of the recognition of digital persons ("Ethics in Digital Futures"), before asking how digital future people might think about corporate groups ("Corporate Groups in Digital Futures"), and finally asking how we should respond to digital future ethics ("Should We Adopt the Ethics of This Digital Future?").

Ethics in Digital Futures

Extending the ethical domain to include digital persons alongside human persons raises many fascinating challenges (Bostrom 2014; Hanson 2016; Mulgan 2014, 2016b, 2018; Yudkowsky 2008). I focus here on two factors that are directly relevant to our present inquiry.

First, and most obvious, digital future ethics must endorse inclusive criteria for agency and personhood. If our paradigm moral persons include both adult humans and digital persons, then our criteria for both moral agency and personhood must cover both groups. Philosophers immersed in a digital future will regard criteria tailored to human embodiment or physical individuality as obsolete and anthropocentric—no better than long-discredited criteria based on race, gender, or class. (After all, some future philosophers will probably *be* digital persons!) Future philosophers who recognise digital personhood will either agree that digital beings *are* intentional, rational, conscious, and free, or else they will deny that these features are truly essential for either moral agency or personhood. And for those features that they judge to be essential, they will insist on *interpretations* that accommodate digital persons. They will thus reject biological interpretations that emphasise our particular organic substrate in favour of non-biological functional interpretations where what matters is what someone does, not what they are made of.

The second distinctive feature of digital ethics is more surprising. Every digital future shares many distinctive features of the *broken* future. In the first place, any digital future threatens to descend into a particularly unpleasant broken future where resources are insufficient to support all existing digital beings and the price of labour falls far below the cost of keeping any (flesh-and-blood) human labourer alive. This is due to the threat of a digital population explosion. Unlike human beings, whose reproduction is limited by biology, natural resources, and inclination, digital beings can reproduce at will. And they may have strong incentives to do so. For instance, Robin Hanson speculates that, in a competitive market, "emulations'" based on a few thousand "exceptional" humans could both dominate the digital economy and effortlessly outcompete human labour—perhaps by selling short-lived copies that do a full day's work and then expire without enjoying any leisure time (Hanson 2016). Flesh-and-blood humans would be overwhelmed by a population explosion that, from their (comparatively slow) human perspective, would seem more or less instantaneous (Bostrom 2014, pp. 22–51).

Like human beings and corporate groups, *digital* beings could be psychopathic or otherwise morally unreliable. Indeed, this is quite likely, for several reasons: it may be much easier to engineer artificial agents who don't respond to moral reasons than ones who do (Bostrom 2014, pp. 105–114); psychopathic or morally unreliable humans may be more likely to have both the resources and the inclination to upload and multiply themselves; and the uploading process itself might undermine a person's concern for her fellow humans (along with her sanity).

The twin threats of digital population explosion and morally unreliable digital beings exacerbate one another. If digital reproduction is constrained only by internalised moral norms, then a single morally unreliable digital being could very quickly dominate a law-abiding population!

Any stable future digital society must therefore find ways to prevent the emergence and/or proliferation of morally unreliable digital agents, without eliminating digital persons altogether. This task raises much-discussed practical difficulties (e.g. Bostrom 2014, Chap. 9; Chalmers 2010; Yudkowsky 2008). However, I set these aside here and assume that future people will successfully eliminate (or neutralise) morally unreliable digital beings. This daunting task will be much easier in an *already* broken future where regulation is more widespread, privacy is reduced, essential resources are scarcer, collective scrutiny is the norm, and techniques for reliably predicting, influencing, or controlling *human* motivations are more reliable than anything available today.

Any digital future that escapes a digital population explosion is therefore a place where digital reproduction is very tightly constrained by both external sanctions and internalised moral norms. And those sanctions and norms must constrain all human persons as well as digital ones. (A single rogue human could quickly produce a vast number of morally unreliable digital beings.) Any digital society that endures for any length of time will restrict the rights and freedoms of all persons to a greater degree than any contemporary affluent liberal society. In the digital future, no one enjoys the right to create *new* agents who are not morally reliable. *Morally reliable* digital beings themselves will endorse this constraint, because morally unreliable digital

Footnote 12 (continued)

regard digital intelligence as a metaphysical impossibility (e.g. Searle 1982, 1997)? I would argue not. I need only claim that my digital future is *epistemically* possible, not that it is *metaphysically* possible. And it clearly *is* epistemically possible, precisely because no one can be sure that Searle is not wrong! I return to related issues in "Should We Adopt the Ethics of This Digital Future?" section.

beings threaten other digital persons as well as flesh-and-blood humans.

Many distinctive features of broken world ethics thus re-emerge in *every* digital world, even one that didn't initially emerge from an already broken future. Freedom is tightly constrained, essential resources are (potentially) extremely scarce, individual self-interest is trumped by collective security and survival, and so on.

The brokenness of the digital future is important for two reasons. First, it raises the importance of broken future ethics, by demonstrating that the set of broken possible futures is much larger than it seems. (We should think of the digital future as an *instance* of the broken future rather than an *alternative* to it!) Second, and more optimistically, future digital philosophers can help themselves to the ethical lessons and resources of the broken future, especially in their dealings with corporate groups.

Corporate Groups in Digital Futures

A central challenge for digital future ethics is to prevent the emergence and proliferation of morally unreliable corporate groups, especially those whose members are themselves digital beings. How might this be achieved?

If every digital future is also a broken one, then digital future philosophers can borrow all the arguments of "Broken Futures" section. Individualism and non-moralised collectivism are thus already on the back foot. I shall now argue that there are additional reasons why individualism is not tenable in a digital future, and why non-moralised collectivism could be disastrous. Only moralised collectivism can meet the threat of the inhuman corporate future.

Digital beings are not corporate groups. However, once digital persons are recognised, then corporate moral agency is much harder to deny. The hardest criteria for *corporate groups* to satisfy are also those that are least friendly to *digital beings*—such as embodiment or ensoulment. Furthermore, I argued in "Ethics in Digital Futures" section that interpretations of intentionality and rationality based on functional roles are friendlier to digital persons than interpretations based on features peculiar to organic humans. And we saw in "Defending Moderate Collectivism" section that the former are also the interpretations that are friendlier to corporate groups! If future philosophers endorse criteria that include digital persons rather than ones moulded to the distinctive features of *human* persons, then they are much more likely to also recognise the moral agency of corporate groups.

Indeed, in the current debate about corporate groups, philosophers on both sides often use robots or computers as an analogy—either arguing that corporate groups *should* count as moral agents because robots would, or that *neither* should count. (Hess 2014; Pettit 2014, p. 1645; Rovane 2014, p. 1670; Velasquez 2003) In a real world populated by digital beings, the latter option is likely to be untenable.

Velasquez (2003) is a very good case in point. His defence of individualism explicitly cites John Searle's account of consciousness as an emergent feature specific to human biological embodiment. And Searle himself uses that account to deny that artificial intelligences could possibly possess consciousness, understanding, or agency (e.g. Searle 1982, 1997). Philosophers in my digital future will dismiss Searle's position as simple anti-digital prejudice.

I conclude that digital future philosophers will look much less favourably on individualism than contemporary philosophers. Having rejected individualism, future philosophers will almost certainly also go beyond *moderate* collectivism. In my digital future, some digital beings are recognised, not merely as agents, but also as full moral *persons*. This opens the door to analogous extremism about corporate groups. If there is a significant boundary between moral agency and personhood, there will inevitably be digital beings on both sides of that boundary. The most intricate and sophisticated *corporate groups* will then claim recognition as moral persons alongside similarly complex digital beings.

In our unbroken non-digital present, individualists present the slide from moderate collectivism to extreme collectivism as an argument against collectivism. In a digital future, collectivists will embrace that slippery slope. Perhaps extreme collectivism will become the default position!

If future philosophers embrace extreme collectivism, then they will have to recognise *some* corporate groups as moral persons whose interests must be counted alongside those of biological humans and digital persons. (Extreme collectivists thus cannot be normative individualists in List and Pettit's sense.) This raises the worry that digital and corporate interests will swamp human ones. One response is to insist on individualism and strive to ensure that corporate groups are *never* recognised as persons with morally significant interests. I have argued in this section that this response is not temporally robust, because it cannot withstand the social pressures that are likely to emerge in possible futures containing *digital persons*. A more robust solution would recognise that the real threat comes, not from corporate groups per se, but only from *morally unreliable* ones, and then seek to ensure that *their* interests are never regarded as morally significant.

To be tenable in a given future, an account of corporate groups must both (a) make sense against the background of accepted social realities and (b) help future people to avoid the worst threats. Non-moralised extreme collectivism is extremely dangerous in precisely those digital futures where moderate collectivism is least credible—namely those where digital persons are ubiquitous—because it threatens to usher in the inhuman corporate future, which is perhaps the worst possible future of all. (If unreliable corporate groups are

recognised as moral *persons*, then they could easily proliferate very quickly!)

I argued earlier that any enduring digital society must limit *digital* reproduction and eliminate morally unreliable *digital* beings. While this *reduces* the risk of an inhuman corporate future, it doesn't *eliminate* it. It is not sufficient to ensure that all *individuals* are morally reliable, because non-psychotic individuals can together constitute a psychopathic corporate group. (At present, after all, a psychopathic business entity might be composed entirely of non-psychopathic human individuals.) Digital future people will therefore insist on *moralised* collectivism, where all new corporate groups must also pass their own *Moral Turing Test*. For reasons discussed in "Corporate Groups in Broken Futures" section, the moral resources of broken future ethics, which digital future ethics shares, make this task much less daunting than it would be today.

Should We Adopt the Ethics of This Digital Future?

I have argued that future people who recognise (some) digital beings as moral persons need an account of corporate groups that is consistent with that recognition. If one recognises (some) digital persons, then one cannot consistently fail to recognise (some) corporate groups as (at least) moral agents. Therefore, digital future people will recognise some corporate groups as moral agents, and individualism is not tenable in digital futures.

I now argue that, because *we* need a temporally robust account of corporate agency that will enable our descendants to make sense of their social world, individualism is also not tenable *for us now*.

Whether or not they were persuaded by my arguments in "Corporate Groups in Digital Futures" section, many individualists will reject this final step. Suppose they grant, for the sake of argument, that humans in digital futures would *believe* that some digital beings are persons, and also that this belief would lead them to reject individualism about corporate groups. Individualists will insist that this merely illustrates the obvious point that future people might be radically *mistaken* about metaphysical facts, and that false beliefs about digital personhood might infect their beliefs about corporate groups. So what? Facts about future beliefs provide no reason to change our present beliefs. What matters is not what imaginary future philosophers might *think* of Searle's argument, but whether he is *correct*.

My individualist opponent here is a *realist* who insists that our philosophical goal is to map independent metaphysical reality. The most robust (if rather unambitious) reply is that, while my argument may not persuade realists, it should persuade those pragmatists or consequentialists who favour individualism (at least in part) because they believe it offers the best hope of meeting future challenges. And, as I argued in "Embracing Broken Future Ethics" section, the prospect of broken futures should make pragmatists and consequentialists of us all!

More ambitiously, I believe that reflection on digital futures should give many *realists* good reason to question individualism. Consider a contemporary realist who concedes that moral agency is a function of organisational complexity rather than biological substrate, but who embraces individualism because she thinks no existing corporate group possesses anywhere near the same degree of organisational complexity as a human brain. This realist might agree that *future* corporate groups consisting of vast numbers of interlinked digital beings could very well qualify as moral agents or even persons. She would then recognise that, despite her present commitment to it, individualism is not temporally robust.

This highlights an important respect in which my defence of extreme collectivism is less radical than it may appear. I have argued that any temporally robust ethic must be both extremely inclusive and extremely collectivist. This means that it must acknowledge the *possibility* that *future* digital beings or corporate groups will be full moral persons, and deny that there is any *in principle* objection to digital or corporate personhood. But this is consistent with maintaining both that no *extant* digital entity or corporate group is (even) a moral agent and also that it is not inevitable that digital or corporate persons will *ever* emerge.

Conclusion

Our search for a temporally robust account of corporate groups has reached a surprising conclusion. The only temporally robust account is *moralised extreme collectivism*, where suitably sophisticated morally reliable corporate groups (and only those corporate groups) are recognised as both responsible agents and full moral persons. This account alone is both intelligible and not-too-dangerous across all the credible futures we have explored. We should therefore bequeath social institutions and moral codes that leave open the possibility of extreme collectivism and allow for full digital and corporate personhood, but also rule out *any* recognition of the agency of morally unreliable corporate groups. We need to learn these lessons from credible futures *now*, because the inhuman corporate future can only be avoided by reorienting our thinking before morally unreliable digital beings emerge and start populating the world with corporate groups whose priorities are entirely divorced from any human concerns.

Acknowledgements I am very grateful to Samuel Mansell and two anonymous reviewers for very detailed and insightful comments on earlier drafts of this paper.

Compliance with Ethical Standards

Conflict of interest Tim Mulgan declares that he has no conflict of interest.

Research Involving Human and Animal Participants This article does not contain any studies with human participants or animals performed by any of the authors.

References

Arnold, D. (2016). Corporate responsibility, democracy, and climate change. *Midwest Studies in Philosophy, 40*, 252–261.

Ashman, I., & Winstanley, D. (2007). For or against corporate identity? Personification and the problem of moral agency. *Journal of Business Ethics, 76*, 83–95.

Bakan, J. (2004). *The corporation. The pathological pursuit of profit and power*. London: Constable.

Bevan, D., & Corvellec, H. (2007). The impossibility of a corporate ethics: For a Levinasian approach to managerial ethics. *Business Ethics: A European Review, 16*, 208–219.

Blackford, R., & Broderick, D. (2014). *Intelligence unbound: The future of uploaded and machine minds*. Oxford: Wiley-Blackwell.

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

Bratman, M. (2000). Valuing and the will. *Philosophical Perspectives: Action and Freedom, 14*, 249–265.

Chalmers, D. (2010). The singularity: A philosophical analysis. *Journal of Consciousness Studies, 17*, 7–65.

Copp, D. (2007). The collective moral autonomy thesis. *Journal of Social Philosophy, 38,* 369–388.

Donaldson, T. (1982). *Corporations and morality*. Englewood Cliffs, NJ: Prentice-Hall.

Dubbink, W., & Smith, J. (2011). A political account of corporate moral responsibility. *Ethical Theory and Moral Practice, 14*, 223–246.

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford: Oxford University Press.

French, P. (1979). The corporation as a moral person. *American Philosophical Quarterly, 16*, 207–215.

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

Goodpaster, K., & Matthews, J. (1982). Can a corporation have a conscience? *Harvard Business Review, 60*, 132–141.

Hanson, R. (2016). *The Age of EM*. Oxford: Oxford University Press.

Hasnas, J. (2013). Whither stakeholder theory? A guide for the perplexed revisited. *Journal of Business Ethics, 112*, 47–57.

Hauskeller, M. (2013). *Better humans? Understanding the enhancement project*. Durham: Acumen.

Heede, R. (2014). Tracing anthropogenic carbon dioxide and methane emissions to fossil fuel and cement producers, 1854-2010. *Climatic Change, 122*, 229–241.

Hess, K. M. (2013). If you tickle us … how corporations can be moral agents without being persons. *Journal of Value Inquiry, 47*, 319–335.

Hess, K. M. (2014). The free will of corporations (and other collectives). *Philosophical Studies, 168*, 241–260.

Hindriks, F. (2014). How autonomous are collective agents? Corporate rights and normative individualism. *Erkenntnis, 79*, 1565–1585.

Huebner, B. (2014). *Macrocognition*. Oxford: Oxford University Press.

Hussain, W., & Moriarty, J. (2016). Accountable to whom? Rethinking the role of corporations in political CSR. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-016-3027-8.

Kusch, M. (2014). The metaphysics and politics of corporate personhood. *Erkenntnis, 79*, 1587–1600.

List, C. (2016). What is it like to be a group agent? *Nous*. https://doi.org/10.1111/nous.12162.

List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford: Oxford University Press.

Malatesti, L. (2010). Moral understanding in the psychopath. *Synthesis Philosophica, 24*(2), 337–348.

Miller, S., & Makela, P. (2005). The collectivist approach to collective moral responsibility. *Metaphilosophy, 36*, 634–651.

Mulgan, T. (2011). *Ethics for a broken world: reimagining philosophy after catastrophe*. Durham: Acumen.

Mulgan, T. (2014). Ethics for possible futures. *Proceedings of the Aristotelian Society, 114,* 57–73.

Mulgan, T. (2015a). Theory and intuition in a broken world. In S. -G. Chappell (Ed.), *Intuition, Theory, and Anti-Theory* (pp. 141–166). Oxford: Oxford University Press.

Mulgan, T. (2015b). Utilitarianism for a broken world. *Utilitas, 27,* 92–114.

Mulgan, T. (2016a). Answering to future people: Responsibility for climate cange in a breaking world. *Journal of Applied Philosophy*. https://doi.org/10.1111/japp.12222.

Mulgan, T. (2016b). Theorising about justice for a broken world. In K. Watene & J. Drydyk (Eds.), *Theorizing Justice: Crucial Insights and Future Directions* (pp. 15–33). London: Rowman and Littlefield.

Mulgan, T. (2017). How should utilitarians think about the future? *Canadian Journal of Philosophy, 47,* 290–312.

Mulgan, T. (2018). Moral imaginativeness, moral creativity, and possible futures. In B. Gaut & M. Kieran (Eds.), *Creativity and Philosophy* (pp. 350–368). New York: Routledge.

Narveson, J. (2002). Collective responsibility. *The Journal of Ethics, 6*, 179–198.

Oreskes, N., & Conway, E. (2010). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. New York: Bloomsbury.

Pettit, P. (1993). *The common mind: An essay on psychology, society and politics*. Oxford: Oxford University Press.

Pettit, P. (2007). Joining the dots. In G. Brennan (Ed.), *Common minds: Themes from the philosophy of Philip Pettit* (pp. 215–344). Oxford: Oxford University Press.

Pettit, P. (2014). Group agents are not expressive, pragmatic or theoretical fictions. *Erkenntnis, 79*, 1641–1662.

Rawls, J. (1971). *A theory of justice*. Cambridge: Harvard University Press.

Rovane, C. (2014). Group agency and individualism. *Erkenntnis, 79*, 1663–1684.

Scherer, A. G., & Palazzo, G. (2011). The new political role of business in a globalized world: A review of a new perspective on CSR and its implications for the firm, governance, and democracy. *Journal of Management Studies, 48*, 899–931.

Schwitzgebel, E. (2015). If materialism is true, the United States is probably conscious. *Philosophical Studies, 172*, 1697–1721.

Searle, J. (1982). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*, 417–457.

Searle, J. (1997). *The mystery of consciousness*. London: Granta.

Theiner, G. (2014). A beginner's guide to group minds. In M. Sprevak & J. Kallestrup (Eds.), *New waves in the philosophy of mind* (pp. 301–322). London: Palgrave.

Tollefsen, D. (2003). Participant reactive attitudes and collective responsibility. *Philosophical Explorations, 6*, 218–234.

Tollefsen, D. (2015). *Groups as agents*. Cambridge: Polity.

Tuomela, R. (2013). *Social ontology: Collective intentionality and group agents*. Oxford: Oxford University Press.

Velasquez, M. (1983). Why corporations are not morally responsible for anything they do. *Business and Professional Ethics Journal, 2*(3), 1–18.

Velasquez, M. (2003). Debunking corporate moral responsibility. *Business ethics quarterly, 13*, 545–546.

Werhane, P. (1985). *Persons, rights, and corporations*. Englewood Cliffs, NJ: Prentice-Hall.

Yudkowsky, E. (2008). Artificial intelligence as a positive and negative factor in global risk. In N. Bostrom & M. Cirkovic (Eds.), *Global catastrophic risks* (pp. 308–345). Oxford: Oxford University Press.