Ethics and Humanity in (Post-)Kantian Philosophy

Jack Wearing
St John's College, University of Oxford

Submitted for the degree of DPhil in Philosophy

Contents

Abstract

Bringing together work in ethics, political philosophy, and history of philosophy, the thesis examines the normative implications of the concept of *humanity*, arguing for a critical, egalitarian humanism that opposes hierarchies of moral standing among human beings while acknowledging our obligations to non-human animals. In the Introduction, I argue for a broad understanding of 'humanism' that captures a variety of different views, each with a different conception of the normative relevance of the notion of the human being. I then elucidate two problems for historically influential forms of humanism, namely that they have instituted unacceptable *hierarchies* among human beings and perpetuated a damaging *anthropocentrism*. I argue that instead of embracing anti-humanism, we should reform or rethink humanism in such a way that it avoids these difficulties. I then critically examine ethical conceptions of 'the human' in the work of four philosophers from apparently disparate traditions – Immanuel Kant, Martin Heidegger, Judith Butler, and Ludwig Wittgenstein. I argue that while neither Kantian humanism nor Heidegger's anti-humanism can convincingly overcome the two problems explained in the Introduction, resources from Butler and Wittgenstein's work can provide a new basis for a more attractive form of humanism that gives central place to the notion of bodily vulnerability and the need to acknowledge it.

Word Count: 75,046 words

I propose to start, and to end, with the question of the human (as if there were any other way for us to start or end!)

Judith Butler,
Precarious Life

We may not know what absolute good is or the absolute norm, we may not even know what man is or the human or humanity – but what the inhuman is we know very well indeed. I would say that the place of moral philosophy today lies more in the concrete denunciation of the inhuman, than in vague and abstract attempts to situate man in his existence.

Theodor W. Adorno,
Problems of Moral Philosophy

Introduction

I. The Sense(s) of 'Humanism'

"How can we restore meaning to the word 'humanism'?" (*LH* 241)[1]

This is one of several questions put to Martin Heidegger in the mid-1940s by the French philosopher Jean Beaufret. Together with a public talk by Jean-Paul Sartre – later published as *Existentialism Is a Humanism* – Beaufret's correspondence provided the occasion for Heidegger's *Letter on "Humanism"*. Characteristically, Heidegger's response to this question in the *Letter* was to bring out the assumptions underlying it, and to raise further questions about their legitimacy: "This question stems from your intention to retain the word 'humanism.' I wonder whether that is necessary"; "Your question [...] also contains an admission that this word has lost its meaning" (*LH* 241, 262). Beaufret's question, then, makes the following two presuppositions: that 'humanism' has in some sense lost its meaning, and that we should *want* to restore it, so as to retain the word and presumably also a version of the project it once denoted. Heidegger is clear that when he says the word has 'lost its meaning', he does not mean that it is "wholly without meaning and a mere *flatus vocis*" – an empty sound to which

[1] I use the following abbreviations for selected works by Martin Heidegger: *BT* for *Being and Time* (1962); *QCT* for 'The Question Concerning Technology' (1977); *LH* for 'Letter on "Humanism"', in his *Pathmarks* [Ga9] (1998); *OM* for 'Overcoming Metaphysics' (1993a); and *OAG* for 'Only a God Can Save Us' (1993b). In accordance with standard practice in Heidegger scholarship, when citing *Being and Time* I give the pagination of the English translation followed by the 'H' numbers, which are printed in the margins and which correspond to the pagination of the seventh German edition of *Sein und Zeit* (see *BT*, 'Translators' Preface', 13). Other references to volumes from Heidegger's *Gesamtausgabe* are given by 'Ga number', followed by the page number of the English translations cited in the bibliography and, in the case of citations of *Contributions to Philosophy (Of the Event)*, by the relevant section number.

nothing corresponds in reality (*LH* 262). Nor is the point that the word has fallen out of usage. After all, Sartre was a self-proclaimed humanist, and provoked Heidegger's ire by claiming that 'existentialism' – under which title he explicitly included Heidegger's own early work – is a form of humanism.[2]

So if the word is not *wholly* meaningless, if it is not the case that 'humanism' is a mere sound without a referent, what might the loss of meaning that Heidegger identifies consist in? Perhaps his claim is akin to a more familiar – though controversial – suggestion from philosophers such as Elizabeth Anscombe and Alasdair MacIntyre that although we continue to use certain forms of expression, especially in our moral discourse, these forms have become in an important sense obsolete. For Anscombe, the notion of specifically *moral* obligation, the *moral* ought, and related concepts, "ought to be jettisoned if this is psychologically possible; because they are survivals, or derivatives from survivals, from an earlier conception of ethics which no longer generally survives, and are only harmful without it."[3] Although these concepts survive – "with a sort of atmosphere clinging to them", as Cora Diamond has put it – they "require for their

[2] See Sartre (2007: 20, 22). See Chapter Two, Section I below for a brief account of why Heidegger refused the labels of 'existentialism' and 'philosophy of existence', despite the apparent centrality of 'existential' themes in his major early work, *Being and Time*.

[3] Anscombe (1958: 1). The 'earlier conception of ethics' that she has in mind is a "*law* conception of ethics", a conception which Anscombe claims "it is not possible to have [...] unless you believe in God as a law-giver". Given the long dominance of this conception under Christianity, Anscombe writes that it is "a natural result" that the concepts of "'obligation', of being bound or required as by a law, should remain though they had lost their root" (1958: 5-6). I cannot evaluate Anscombe's argument here, though it is perhaps worth noting that Heidegger also appeals to the metaphor of the rootedness – or rather, the uprootedness – of our concepts. See Ga65: 214, §150: "all discussion of *existentia* and *essentia* remain [...] an empty wrangling over uprooted concepts"; see further Ga65: 276, §225 for another use of this metaphor. This point foreshadows Heidegger's later criticism of Sartre, namely that Sartre's motto 'existence precedes essence' is merely the reversal of a metaphysical formula, and does not achieve an understanding of 'existence' – or as Heidegger will later call it, 'ek-sistence' – that goes beyond the metaphysical tradition (*LH* 250). I will have more to say about this criticism in Chapter Two.

content or intelligibility background conditions which are no longer fulfilled." Thus, "we go on using the old words, but the old words can no longer carry their old significance."[4]

If an interpretation along these lines is correct, then Heidegger's point is not just that a word, 'humanism', no longer has a literal meaning, but that the project that this word once denoted has lost its significance or intelligibility in a broader sense – what we could perhaps call, tendentiously but I hope suggestively, an 'existential' sense. If so, Beaufret's question concerns not just the meaning of 'humanism', but the importance or even the credibility of humanism: what meaning can humanism have for us, now?

Heidegger accepted Beaufret's first presupposition – that 'humanism', and therefore humanism, had in some sense become meaningless in his time – but rejected the second. What is needed is not, he thought, the restoration of sense to 'humanism', but a philosophy that went beyond humanism in any traditional sense. This is because – for complex reasons that I shall go into later – Heidegger claimed that "the essence of humanism is metaphysical" (*LH* 262). Humanism has uncritically taken over a conception of the human essence from the metaphysical tradition. While "not false", this conception nonetheless remains "conditioned by metaphysics" in a problematic sense: "The first humanism, Roman humanism, and every kind that has emerged from that time to the present, has presupposed the most universal 'essence' of the human

[4] See Diamond (1988: 257). Compare here Heidegger's claim that our situation is characterized by "the exhaustion of every basic word and the destruction of the genuine relation to words" (Ga65: 5 [Unnumbered Preface]). What the restoration of a 'genuine relation to words' might amount to is not clear, though his reinterpretation in the *Letter* of 'ethics' in light of what he claims to be the "basic meaning" of the Greek word '*ethos*' – "abode, dwelling place" – perhaps provides an illustration of what he has in mind (see *LH* 269-71).

being to be obvious. The human being is considered to be an *animal rationale*" (*LH* 245-6).

What is wrong with this metaphysical conception of our essence? Heidegger offers a grand narrative of the development and "completion" of metaphysics – part of his "history of Being" – according to which the final, completed stage of metaphysics is "technology" (OM 67, 75).[5] For Heidegger, the term 'technology' does not refer to – or does not merely refer to – machines and technological innovations of human design, but to a metaphysical outlook according to which entities increasingly appear to us only as resources and means.[6] What worries Heidegger most of all about this development is the way in which it conceals other possible ways of encountering entities, other understandings of the Being of those entities, and ultimately the question of Being itself. Metaphysics, technology, and indeed humanism thus on the one hand enjoy an unprecedented dominance in modernity, but on the other hand represent, in Heidegger's view, a disastrous dead end. Thus, he asks: "Should we still keep the name 'humanism' [...] just so that by sharing in the use of the name we might perhaps swim in the predominant currents, stifled in metaphysical subjectivism and submerged in oblivion of Being?" (*LH* 262) The alternative, according to Heidegger, is to cultivate a new kind of thinking that can return us to what he takes to be our genuine essence: to be "called by Being itself into the preservation of Being's truth", and thereby to overcome the reductive understanding of Being involved in metaphysical, technological thinking (*LH* 260; see also 247).

[5] I have systematically capitalized 'Being' where it translates '*Sein*' in order to avoid confusion with 'being' in the sense of 'entity'.
[6] I elaborate on this claim in Section III below.

Heidegger's own reasons for rejecting humanism were, to put it mildly, idiosyncratic. They are inseparable from other substantial commitments of his later philosophy, which are in turn likely to strike many readers as highly obscure.[7] Nonetheless, despite the idiosyncrasy of his 'anti-humanism', Heidegger's *Letter* has been viewed as inaugurating a dispute about the use of the concept of the 'human' in ethics and philosophy more generally, especially in France in the post-war period. Thomas Baldwin describes this so-called 'Humanism Debate' as "an extended critical discussion of the relationship between ethical values and human nature."[8]

However, this characterisation is in two respects somewhat misleading. The first is that it is something of an oversimplification to talk as if there has been one unified discussion about 'humanism', even if we restrict our view to the French context. As is evident from Baldwin's own overview, concerns about humanism and the ethical status of the concept of the human in 'Continental' philosophy arose from a variety of sources. Some contributions to these debates are evidently responsive to Heidegger's particular concerns, others less so.[9] We may also take a broader view, and see debates about humanism arising on both sides of the familiar – and familiarly problematic – divide between 'Continental' and 'analytic' philosophy. From this broader perspective, the 'debate' will look still less unified. While there are important points of continuity

[7] See further Section III below; see also Chapter Two for a more extensive reconstruction and critical discussion of Heidegger's 'anti-humanism'.
[8] Baldwin (2007: 672). See also Phillips (2015: 14ff.).
[9] For an example of the former, see Derrida (1969); see also Baldwin's discussion of Derrida in (2007: 703ff.). For two rather different examples of the latter, see Althusser (2005: 219-248); Foucault (2002). See especially the provocative claim with which Foucault concludes this book: "man is an invention of recent date. And one perhaps nearing its end"; if our fundamental arrangements of knowledge were "to crumble [...] then one can certainly wager that man would be erased, like a face drawn in sand at the edge of the sea" (2002: 422).

and shared concern among the various disputants in these discussions, the points of divergence are also worth noting. A charge levelled against 'humanistic' philosophers in the analytic tradition, most famously by Peter Singer, is that of 'speciesism'.[10] This label is intended to suggest, on the model of 'racism' and 'sexism', that to accord to human beings a special moral status merely *qua* human beings – which Singer takes to mean *qua* members of the species *homo sapiens* – is to exhibit a prejudice or bias against other species: the 'human prejudice', as Bernard Williams has called it, while disputing that it is really any such thing.[11] This complaint has something in common with a charge made by some writers from the Continental tradition, albeit under the rubric of anthropocentrism rather than speciesism.[12] In both cases, the broad claim is that humanism gives a problematic centrality to human beings' concerns, to the exclusion of non-human animals and perhaps non-human nature more generally. Nonetheless, there will be important differences among different thinkers concerning the grounds for the respective charges of anthropocentrism and speciesism, and not all variations on these charges will be equally plausible.[13]

---

[10] See e.g. Singer (2011), Chapter 3; (2009a).

[11] See Williams, (2006: 135-152). Singer's response arguably misconstrues the central points of Williams' essay; see Singer (2009c). See also Diamond (2018) for a useful clarification and defence of some of Williams' arguments.

[12] See e.g. Braidotti (2013), Chapter 2; Derrida (2008: 113, 127). Interestingly, and perhaps surprisingly, Derrida approvingly cites the same quotation from Jeremy Bentham that Singer appeals to in several of his arguments against speciesism: "The question is not, Can they *reason*? nor Can they *talk*? but, *can they suffer*?" See Singer (2011: 50), and Derrida, (2008: 27, 103). Derrida describes Bentham's question as simple but profound, and views it as calling into question the priority given to language and the *logos* in discussions of 'the human' and non-human animals. Singer commends Bentham for emphasizing the relevance of suffering to the determination of moral status, but criticizes him on the grounds that, on Singer's own view, intellectual capacities should also be taken to be relevant.

[13] I shall return to this point in Section III below, arguing that the most plausible worries about the exclusionary character of certain forms of humanism also warrant a rejection of Singer's brand of utilitarianism.

The second respect in which Baldwin's characterisation of these debates might mislead is his mention of 'human nature'. This description of the content of the discussion might make it seem too narrow in scope: the phrase 'human nature' often connotes the idea of a substantive conception either of what human beings are like – what motivates them and explains their behaviour – or of what they ought to be like. (The notion of '*Homo economicus*' would be one – implausible – example of such a conception.) To the extent that the 'humanism debates' touch on concerns about the relationship between 'ethical values' and human nature in this sense, they have not generally taken the form of first-order debates about this or that substantive conception of what humans are like, but have rather brought into question the very notion of human nature. Sartre claimed that "there is no human nature, because there is no God to have a conception of it."[14] Other philosophers of a more historicist bent, such as Foucault and some writers in the Marxist tradition, have questioned whether any substantial, transhistorical account of human nature can be given.[15]

In this sense, the 'humanism debates' do concern the use of the notion of human nature, but they also have a broader scope. It will be more useful to think of these discussions as concerning the ethical import of the concept of 'the human' and 'humanity' more generally. What is in question is not only whether there is such a thing as 'human nature', but also whether we can talk – *à la* Sartre – of a universally shared *condition humaine*: whether it is possible to provide a philosophical anthropology or an ontology of the human that could claim explanatory priority over any empirical study of 'human nature'. Further, there is the question of what relevance such accounts

[14] Sartre (2007: 28).
[15] See the citations in fn. 8 above.

would have in relation to normative questions. By what right can we talk, in universal terms, about human flourishing, or even human rights? Do we need a philosophical anthropology or a substantive account of human nature in order to support these notions? Finally, are human beings as such of central concern to ethics at all?

My aim in this thesis will not be to resolve all of these disputes. Rather, I hope to clarify what is at stake in the 'humanism debate'; to bring work from both sides of the analytic/Continental divide into dialogue in order to do so; and to address some central arguments regarding the ethical implications of the concept of *the human*. In doing so, I hope to bring some clarity to a general question that unifies several of the more specific questions I just listed: what role should the concept of *humanity* play in ethics?[16] For my purposes here, I shall use 'humanism' in a broad sense, to refer to views that take 'human being' to be an important ethical concept. In one sense, this provides a stipulative – and therefore trivial – answer to Beaufret's question for Heidegger, with which I began. In another sense, however, that question remains a useful starting point. For we may now reinterpret its underlying line of thought as follows: What should 'humanism' come to mean? What kind of humanism could play a meaningful role in our ethical lives today? Is it necessary to retain this word, and the project it denotes? Or, on the contrary, should we agree with Heidegger – whether for his reasons, or different ones – that there is something *damaging* about humanism (*LH* 241), something ethically problematic?

[16] In contemporary jargon, it would not be inaccurate to call this an enquiry in 'conceptual ethics', focusing on the ethics of the use of the concepts of *humanity* and *human being*. For an attempt to unify this field of normative inquiry, see Burgess and Plunkett (2013).

Of course, I have posed that last question too simplistically: humanism is not just *one* project, and in fact a great variety of ethical positions and projects might warrant that title. As Bernard Williams notes, the term is now most prominently ascribed to – or used as a self-ascription by proponents of – a certain species of sometimes militant atheism.[17] In its most Polyanna-ish forms, this kind of view may be coupled with startlingly strong claims about historical progress, Enlightenment, and the great powers of human reason – a coupling emphasised, for instance, in the subtitle of Steven Pinker's recent book, *Enlightenment Now: The Case for Reason, Science, Humanism and Progress*. Pinker declares that "The Enlightenment has *worked*," that we are living in the most prosperous and peaceful era in humanity's history – vast global inequality, the atrocities of the twentieth century, the recent resurgence of regressive populist movements across the world, and the fact that humanity has never been closer to bringing about its own extinction than it has in the past century all apparently notwithstanding.[18] (In a quote on the front cover, Bill Gates, currently the fifth richest man in the world, calls it his "new favourite book of all time.") But humanism need be neither secularist nor especially optimistic. Heidegger explicitly argues that "Christianity

[17] Williams (2006: 135); see also Phillips (2015: 15, 107).

[18] Pinker (2019: 6) and *passim*. Robin Celikates has – not unfairly – characterised Pinker's conception of progress as a form of triumphalism; see Celikates (2018: 144, 147). See also Manne (2016) for criticism of contemporary forms of humanism, including that of Pinker. One of Manne's central points is that there has been an overemphasis on 'dehumanisation' as an explanation for violence. While this claim is plausible as far as it goes, it does not of course follow that the concept of dehumanisation has *no* important role to play in ethics; see especially Section IV below, as well as Chapters One and Three. It is also worth noting that Manne is unfortunately overhasty in lumping the views of philosophers such as Cora Diamond together with more Pinkerian versions of humanism. Part of my aim in this introduction, and the thesis as a whole, is to show that such overly general dismissals of 'humanism' are unlikely to be very useful. Finally, in Manne's use of the term, 'humanism' refers to an explanatory view in moral psychology (2016: 390), whereas in my usage it refers to a set of normative views. To that extent, Manne and I are concerned with different (albeit overlapping) phenomena. With that said, for reasons that will become apparent I am in agreement with Manne's claim that humanism is not necessarily "committed to morally suspect claims about the superior value or greater rights of human beings relative to other creatures" (2016: 391, fn. 5).

too is a humanism, in that according to its teaching everything depends on human salvation" (*LH* 245). Williams suggests something similar in his discussions of Martin Luther; even in his pessimistic vision of human beings as "hideously fallen", a "central belief" held by Luther is that "man's fate is a very special concern to God", and thus "a central concern, period."[19] There are also Marxist forms of humanism, which, while generally committed to secularism, are unsurprisingly much more critical of contemporary socio-political conditions than liberal forms of humanism.[20]

Given this diversity of philosophical and ethical projects that can be understood as humanistic in character, it would seem futile and counterproductive to try to formulate a precise and substantive definition of 'humanism'. We would do better to acknowledge the diversity of possible 'humanisms', and to ask, not whether humanism as such is ethically defensible (or indeed necessary), but rather what kind of humanism might be so. Nonetheless, it will be useful to adopt a loose – and, to as great an extent as possible, neutral[21] – characterisation of humanism for my purposes here.[22] Cora

[19] Williams (2006: 136; see also 147).

[20] See Fromm and Marx (2013) for excerpts from Marx's *Economic and Philosophical Manuscripts* that have greatly influenced the development of Marxist versions of humanism, accompanied by Erich Fromm's lengthy introduction. The question of whether Marx's mature critique of capitalism retains the apparently humanistic views elaborated in this early work remains, however, a matter of controversy within Marxist theory. See Geras (2016) for criticism of Althusser's 'anti-humanist' reading of Marx's later work (cited in fn. 8 above); see also Butler (2019) for a reading that disputes a straightforward imputation of anthropocentrism to Marx's early work.

[21] By a 'neutral' characterisation, I mean primarily one that is not pejorative – see my comments on contemporary uses of 'humanism' and 'anthropocentrism' in Section II below. Compare Kagan (2016: 2), in which Kagan makes a similar move regarding the term 'speciesism': "If we are going to objectively evaluate the charge that speciesism is a mere prejudice and nothing more, we had better not build the claim that it is a prejudice into the very definition of the term." Though I think there is good reason to avoid stripping 'speciesism' of its negative connotations, like Kagan I am unconvinced that it is an apt label for what most of us believe.

[22] Compare Adrian Moore's comments on his broad 'working definition' of 'metaphysics' in Moore (2012: 1): "to have tried to attain substance without controversy would have been

Diamond's description of "the view which Singer attacks and which Williams defends", with some minor modifications, strikes me as a useful starting point. Humanism, on this conception, is the view

> that "human being" is a [normatively] relevant notion – that is, a notion that can appropriately be used (in [normative] thinking) to give a reason why some creature should [or should not] be treated in such-and-such a way.[23]

Allow me to briefly comment on this variation on Diamond's description. First of all, it should be admitted that not all humanists would frame their view in terms of 'giving a reason'. What I mean by this, however, is just that the concept 'human being' is "action-guiding".[24] Proponents of different forms of humanism will disagree about the ways in which this concept should guide our action – about what forms of action count as respectful of human dignity, or about which human rights are basic and universal. They will also disagree about why the concept should guide our action. But they will not disagree that there is *some* substantive relationship between the fact that someone is a human being and what we owe to them – in other words, that someone's humanity has a bearing on how we may and must behave towards them. This broad characterisation is compatible both with views that seek to explain and justify the normative relevance of the concept of the human being (for instance by appeal to substantive normative accounts of human nature, dignity, or flourishing) and with

---

foolhardy", since the nature of metaphysics remains "fiercely contested". Something similar could surely be said of humanism.

[23] Diamond (2018: 381). I have here substituted 'normatively' for 'morally' and 'normative' for 'moral' to obviate a narrow reading of 'moral' that distinguishes it from the political, since I take it that at least some forms of 'humanism' claim overtly political uses for 'the human'. I have also added the phrase 'or should not', primarily because human rights declarations, which are paradigmatically 'humanistic' documents on my conception, standardly frame many of our rights in negative terms.

[24] For a useful discussion of action-guiding concepts, see Moore (2003a: 39-51).

views on which we do not need any further reason or justification to take 'human being' to be a normatively relevant notion, such as that of Bernard Williams. For Williams, that someone is a human being just is a basic reason for us to treat them in certain ways (such as trying to rescue them when they are in immediate danger), and refrain from treating them in other ways (such as torturing them).[25] Anne Phillips endorses a view that is in some ways similar when she writes "I do not [...] see [our] species features as *justifying* our moral status – and this is not because something else justifies it instead, but because justification of that status is not the issue."[26]

Furthermore, this characterisation implies nothing about what 'the humanist' ought to say about our treatment of non-human animals. In particular, it does not rule out the possibility that 'the humanist' might be in a position to recognise strong ethical reasons for objecting to many currently widespread ways of treating non-human animals (for instance killing and eating them). It is simply non-committal on such issues. To say that 'human being' is a normatively relevant notion is not to say that other notions ('sentient', for example) are not also normatively relevant, and to say that human beings are important does not imply that they are *uniquely* important. Some forms of humanism do of course have this implication, but there is little reason to build it into

[25] See Williams (2006: 139-40); Diamond (2018: 380-1). See further Section IV below.

[26] Phillips (2015: 29). It is worth noting here that Phillips actually disavows the label 'humanism', though this is in part because her characterization associates it more narrowly with problematic extant or historically influential forms of humanism (2015: 14ff., 107). In the terms of my broad characterization, however, her account of the politics of the human – which stresses the close connection between the notion of humanity and the idea of equality, and strongly rejects the notion of a hierarchy among human beings – is strongly humanistic (2015: 79, 46). I thus take my dispute with her to be primarily terminological, though I would not claim that there are no good reasons for restricting the scope of 'humanism' in the way that Phillips does.

the very definition of 'humanism'. I shall return to this line of thought frequently in what follows.

With this broad characterisation of 'humanism' in hand, my aims in the remainder of this Introduction are as follows. In Section II, I discuss some important considerations that have been taken as strong objections to humanism. In particular, I explain two ways in which historically influential forms of humanism have been problematically exclusionary: they have relied on or instituted *hierarchical* conceptions of humanity, and have thereby excluded some human beings from moral consideration; and they have been unacceptably *anthropocentric* in their exclusion of the interests of non-human animals and in their legitimation of the domination of non-human nature more generally. I argue that while these forms of exclusion are not part of the definition of humanism, avoiding them is a minimal condition that humanism must meet. No form of humanism that fails to take these concerns seriously will be ethically plausible.

We are therefore faced with a choice between rethinking humanism, or accepting some form of anti-humanism. Thus, in Section III I provide a cursory examination of the prospects for anti-humanism. Because the set of views that might warrant this label is if anything even less unified than the various forms of 'humanism', no one view can be taken as representative, but neither would a complete survey be feasible. For this reason, I restrict my attention to two forms of 'anti-humanism' that have little in common and are in some respects sharply opposed, namely the views of Peter Singer and Martin Heidegger. Despite the stark differences between their views, I suggest that neither Singer nor Heidegger offers an improvement on 'humanism' in relation to the

two concerns detailed in Section II. While this is by no means a reason to think that no form of anti-humanism could adequately respond to those charges, it does suggest that different forms of humanism and anti-humanism alike must be considered on an individual basis, and that a turn away from humanism as such may be premature.

In Section IV, I conclude by providing further support for this last claim. I argue that because the notion of the human is deeply embedded in our linguistic and legal practices and is intimately related to many of our ethical commitments, it is difficult to represent the withdrawal of 'humanistic' normative resources as anything other than a loss, or even a form of regression. This conclusion does not, however, license an uncritical adherence to whichever normative conception of the human happens to come our way. On the contrary: it makes the task of renewing humanism all the more urgent.

II. Two Charges Against Humanism: Hierarchical and Anthropocentric Conceptions

The criticisms of humanism with which I shall primarily be concerned are ethical or normative in character. There are certainly other important objections that have been raised against various forms of humanism, for instance that they have made critical descriptive errors in their accounts of human nature, the human condition, or the human essence. Some of these objections will concern me too, insofar as normative claims are often read off from or given ideological support by such erroneous conceptions. In this section, I will briefly outline two charges levelled against humanism, both of which concern forms of exclusion: the *hierarchical* exclusion of particular groups of human beings and the *anthropocentric* exclusion of non-human

animals (and nature more generally). By 'exclusion' here I primarily mean forms of ethical exclusion, for instance a failure to take seriously the interests of those excluded, or a failure to accord them moral status. I will argue that any plausible form of humanism must avoid these exclusionary consequences. I will therefore return to the two charges introduced here frequently when I give more detailed consideration to various forms of humanism and anti-humanism in the chapters to come.

Specific forms of humanism have typically offered substantive accounts of what it is to be a human being. The moral status of the human being is meant to be explained in terms of a distinguishing feature. To take one philosophically influential example, Kant saw the human being's moral worth as grounded in their rationality.[27] However, if the feature such substantive accounts pick out is not possessed by all human beings – or not believed to be possessed by them – then there is a risk that such conceptions will tacitly institute a hierarchy of moral standing among human beings: a division of status between those who fully exhibit the feature in question, and thus who (are believed to) lack it or exhibit it to a lesser degree. Such divisions are then viewed as normative, as justifying differences in the status and treatment of the human beings on either side of the division. Anne Phillips argues, for instance, that since even in relatively recent centuries "women were widely thought not to exhibit [the] qualities" of rationality and the capacity for normative judgment, "it became possible to set women outside the category of the human, to say that all men have equal rights (because of their common humanity) but mean *only men*."[28]

[27] See Chapter One for extensive discussion of Kant's view.
[28] Phillips (2015: 23).

In the case of such hierarchical conceptions of the human, an ethical exclusion will often be legitimised through a conceptual exclusion: someone is not 'counted' as (fully) human, is viewed as sub-human or less-than-human, and that categorisation underlies a degradation of their moral status. Thus, the suspension of basic rights and the 'justification' for certain forms of violence and oppression has often gone hand in hand with the denial of 'full' human status.[29] In the context of Enlightenment thought, and against the backdrop of colonial domination, substantive accounts of the human essence have sometimes been connected with pernicious stories about human 'development'. Consider, for instance, Hegel's remarks about slavery, which come in the context of his highly Eurocentric story about the historical development of the consciousness of freedom. Hegel tells us that "slavery is unjust in and for itself, for the essence of man is freedom". However, he immediately goes on to qualify this judgment:

> ...but he must first become mature before he can be free. Thus, it is more fitting and correct that slavery should be eliminated gradually than that it should be done away with all at once. [...] In rational states, slavery no longer exists; but before such states have come into being, the authentic Idea is present in some areas of life only as an unfulfilled obligation, in which case slavery is still necessary: for it is a moment in the transition towards a higher stage of development.[30]

Hegel makes these remarks in the context of a highly questionable account of the 'Geographical Basis of World History', where he associates different stages in the consciousness of freedom not just with different eras of history, but with different geographical conditions. The inhabitants of Africa, he claims, "see nothing improper

[29] This is a persistent theme in the work of Judith Butler, which I consider in Chapter Three. Butler links strategies of dehumanization to various forms of contemporary racist violence, as well as to homophobic, transphobic, and gender-based violence.
[30] Hegel (1975: 184).

about" slavery. Indeed, he goes so far as to suggest that "slavery has awakened *more humanity* among the negroes" – that is, has brought them closer to the 'essence of man'.[31] Thus, Hegel ends up endorsing a highly equivocal position. While he recognises the injustice of slavery on the basis of his claims about the human essence, he nonetheless seemingly justifies the perpetuation of that injustice in the case of human beings who have not yet reached 'maturity'. His (barely) implicit account of 'maturity' is also highly racialized. The disturbing conclusion seems to be that slavery is still 'necessary' (and thus vindicated) in Africa as a moment in its 'transition towards a higher stage of development'.

This seems to me to be a relatively straightforward example of a normative philosophical account of the human essence that provides ideological support for the perpetuation of racist violence and injustice against certain groups of human beings. Hegel provides a "civilizational discourse" that legitimises colonial domination by excluding the inhabitants of Africa from 'full' participation in the human essence.[32] Nonetheless, while such examples bring into sharp relief the violent uses to which certain conceptions of the human might be put, they do not necessarily warrant a wholesale rejection of 'humanism' in general. One initial response would be to view the charge of hierarchical exclusion not as a criticism of humanism as such, but as an internal or immanent critique – one that relies on broadly humanist intuitions. After all, the problem with such forms of humanism is that they justify violence *towards*

[31] Hegel (1975: 183).
[32] The quoted phrase is taken from Butler, *PL* 72. I use the following abbreviations for selected works by Judith Butler: *GT* for *Gender Trouble* (2006); *PL* for *Precarious Life* (2004); *FW* for *Frames of War* (2016a); *RVR* for 'Rethinking Vulnerability and Resistance' (2016b); and *FN* for *The Force of Nonviolence* (2020).

human beings. So viewed, the criticism would express a demand for a more inclusive form of humanism – one that does justice to the central humanist commitment that all human beings are ethically significant, simply in virtue of their humanity.

Nonetheless, even if we accept this response, the second charge that I mentioned at the start of this section still remains: namely that humanism has had – and will continue to have – the ideological function of justifying violence against or domination of 'the non-human'. It is at least *prima facie* plausible that the emphasis placed on the importance of human beings might serve to eclipse the needs and interests of other creatures. This is especially true when claims about human worth are supplemented by claims about human superiority: the idea "that humans stand higher in the scheme of things than the highest of the other animals."[33] Thus, certain forms of humanism appear to court accusations of anthropocentrism.[34] We should have serious concerns about the ethical plausibility of forms of humanism that cannot respond to this charge.

In the previous paragraph, I implicitly drew a distinction between 'humanism' and 'anthropocentrism'. However, it should be noted that not all philosophers do so. In the analytic tradition, philosophers sympathetic to a conception of philosophy (or, more

[33] Waldron (2017: 4). Waldron himself uses the term 'human dignity' to refer "to conceptions that attribute a high *and distinctive* status to humans, a status that is supposed to contrast with the moral considerability of non-human animals" (2017: 3-4). Waldron later associates human dignity with "the principle of *distinctive* equality", according to which humans "are one another's equals on a basis that does actually differentiate them from animals", and identifies himself as "a strong believer in distinctive equality" (2017: 31). His account of human equality is thus, on my view, not just humanistic but avowedly anthropocentric.

[34] One historically significant source for such views has been Christianity, or certain interpretations of it. Consider here a famous verse from Genesis, which affirms man's divine likeness while granting humanity 'dominion' over the earth: "And God said, Let us make man in our image, after our likeness: and let them have dominion over the fish of the sea, and over the fowl of the air, and over the cattle, and over all the earth, and over every creeping thing that creepeth upon the earth" (*KJV*, Genesis 1:26).

narrowly, metaphysics) as a humanistic discipline are sometimes willing to use 'anthropocentric' as more or less interchangeable with 'humanistic'. In such contexts, neither word is meant to carry an overtly negative connotation. Adrian Moore, for instance, follows Bernard Williams in defend a conception of philosophy on which the human point of view takes a central place.[35] Philosophers writing in or influenced by the poststucturalist tradition often use *both* terms in a more pejorative sense. Such philosophers also often appear to view the terms as closely connected, if not synonymous.[36] Underlying some of the more pejorative usages there is a somewhat ironic tendency to essentialise: to assume on the basis of the – real and highly significant – exclusions of historically influential forms of humanism that humanism as such is exclusionary or otherwise ethically suspect. However, as I suggested in Section I, we should take the question of whether any *particular* form of humanism is ethically defensible to be genuinely substantive. It therefore should not be decided by fiat, but rather on the basis of an examination of the particular normative content it attributes to the concept of the human.

Nonetheless, it would be useful to retain a more pejorative expression as well, because given the preceding arguments, there are clearly ethical failings of which various forms of humanism have historically been guilty. I therefore suggest that we reserve the term 'anthropocentrism' and its cognates for forms of humanism that are problematically exclusionary of 'the non-human'. While this would be somewhat revisionary, it would not be radically so: the term is already often used in a pejorative sense in philosophical

[35] See Moore (2012: 602-4); (2017).
[36] For discussions that link humanism to anthropocentrism, see e.g. Derrida (2008: 113, 127); Braidotti (2013: 69); [James] Williams (2015: 304). For an overtly critical usage of the term 'humanist', see e.g. Wolfe (2003: 51).

contexts and elsewhere, for instance in discussions about climate change. Furthermore, just as 'speciesism' plays on an association with terms like 'racism' and 'sexism', 'anthropocentrism' can be heard as analogous to terms like 'Eurocentrism'. Just as that term suggests an inordinate focus on Europe to the exclusion of the rest of the world, 'anthropocentrism' suggests an inordinate focus on the human to the exclusion of non-human interests. It is also a more flexible term than 'speciesism', because it does not necessarily suggest a prejudice or discrimination based on a specifically biological determination: it could also denote an exclusion based on a substantive philosophical or ethical conception of human being – the '*anthropos*' of anthropocentrism.

Kant's humanism can provide a brief illustration of this point here.[37] For Kant, using a human being as a mere means is categorically impermissible, for we must respect humanity as an end in itself. But Kant explains the obligation to treat human beings as ends-in-themselves by appealing to the notion of rationality. Human beings must always be treated as ends because – or insofar as – they are rational agents, *persons*, capable of freely setting their own ends. This is what constitutes their inner worth. But Kant's moral theory – at least as he himself formulates it – leaves no room for the idea that non-human animals might also have moral status, inner worth – might also be regarded as ends-in-themselves. Insofar as non-human animals are beings without reason, Kant argues that they have the status of *things*, with merely instrumental value. Non-human animals are not the kind of beings to which we could owe duties, or indeed respect, so there can at best be an indirect justification for humane treatment of them. Kant's view is humanistic, insofar as it treats 'human being' and 'humanity' as normatively relevant

[37] I discuss Kant's views in much greater detail in Chapter One.

notions. But, because of the particular account it gives of the moral status of human beings, it is also deeply anthropocentric in the pejorative sense.

In my proposed terminology, then, the difference between humanism and anthropocentrism amounts to a difference between treating humanity as being essentially ethically significant, and doing so *to the exclusion of the interests of non-human beings.* [38] Humanism, in the loose sense defined above, cannot be said to be problematically exclusionary merely by definition, since one can in principle accord to human beings a basic ethical significance, while also taking non-human interests into consideration. Indeed, one can recognise significant and ethically salient asymmetries between humans and non-humans without excluding the latter from moral consideration. Even if many influential forms of humanism have also been anthropocentric, we should not rule out in advance the possibility that humanism can avoid anthropocentrism.

III. Anti-Humanism and its Limits

Given the severe charges raised against extant forms of humanism that I outlined in Section II, we are faced with a choice between two broad categories of response: either to reform or rethink humanism so as to make it responsive to these worries, or to

[38] One complication worth noting here is that anthropocentrism can also sometimes support the institution of hierarchies among human beings, if some of them are regarded as non-human. As James Williams puts it, "Anthropocentrism [...] can justify exclusion and violent behavior toward the nonhuman. Historically, the nonhuman has included groups we now consider fully human" (2015: 305). In such cases, the two potentially problematic dimensions of normative conceptions of the human that we have been considering will be intimately related. Alice Crary makes a related point when she argues that "the denigration of animals [is] internal to much racist rhetoric"; see Crary (2019a: 245). Strategies of dehumanisation thus often rely on "vilifying [...] animal comparisons in discourse intended to degrade groups of human beings" (2019a: 236). The denigration of non-human animals and the degradation of human beings is thus intertwined in such discourse.

abandon it in favour of some form of anti-humanism. I will take 'anti-humanism' to mean the denial of humanism, in the broad sense I introduced in Section I.[39] There can be no question of taking any particular philosopher as a straightforward representative of anti-humanism, since – their rejection of humanism notwithstanding – proponents of anti-humanism may have few shared positive commitments. Nor will it be feasible to offer an exhaustive survey or appraisal of anti-humanistic positions here, and I will not aim to do so. Nonetheless, in this section I will briefly present two anti-humanist views – namely those of Peter Singer and Martin Heidegger – with two aims in mind. The first is to show just how disparate the reasons for rejecting 'humanism' may be. This suggests that there is little point in staging the debate between 'humanism' and 'anti-humanism' at a very high level of abstraction. Specific variations on either broad position must be considered on their own merits. Nonetheless, the second aim will be to cast doubt on the idea that a straightforward denial of humanism is the best response to the two concerns I set out in the previous section. Neither position under consideration succeeds in securing what we might have wanted from humanism, and faulted humanism for failing to provide. None of this is to say that no form of anti-humanism could provide an adequate response to those problems. It is only to present a *prima facie* case that a recoil from humanism may be premature.

[39] There is perhaps something misleading about framing the 'choice' and the alternatives in such stark terms. In particular, one might think that there is a third category of response here, namely to adopt one of the views that fall under the label of 'posthumanism'. Like 'humanism' and 'anti-humanism', 'posthumanism' is by no means a monolithic position. The views that claim that title – or have been subsumed under it – are not homogeneous, though they are typically marked by their interest in an overlapping set of concerns. Chief among these are the possibilities of radical transformations promised – or threatened – by technology, and the concern to 'decentre' the human in various ways, especially in relation to other species and the environment. See Phillips (2015), Chapter 5, for a discussion of three forms of posthumanism; see further Braidotti (2013) for a more partisan introductory overview. For my purposes here, however, we can view extant posthumanist views as collapsing either into 'humanism' or 'anti-humanism' as defined in the main text, depending on the ethical claims they endorse.

Let us first consider Singer's reasons for rejecting humanism. As we saw earlier, Singer's work is strongly associated with the claim that it would be *speciesist* to take 'human being' to be a normatively relevant notion. Singer's argument for this claim proceeds from his commitment to an apparently egalitarian principle: the principle of equal consideration of interests. According to Singer's principle, we must "give equal weight in our moral deliberations to the like interests of all those affected by our actions"; the principle is supposed to guarantee impartiality, since "an interest is an interest, whoever's interest it may be."[40] With this principle in hand, Singer argues that there is an analogy between racist discrimination and our collective treatment of non-human animals. Racists, Singer writes, "violate the principle of equality by giving greater weight to the interests of members of their own race"; similarly, those who give greater weight to the interests of human beings merely in virtue of the fact that they are of the same species – in Singer's view, most of us – can be viewed as speciesists.[41] This, he concludes, is a morally unjustifiable prejudice.

However, matters are more complicated than this simple argument suggests. So far, this is only an argument against the claim of human *superiority*, not against humanism as I have formulated it. For one could take the concept 'human being' to be normatively relevant without also thinking that it can be employed as a tie-breaker whenever we are weighing up the like interests of human beings and non-human animals. (Indeed, one might think that moral deliberation often has nothing to do with weighing up interests.) Thus, one could accept Singer's conclusion – that it would be speciesist to

[40] Singer (2011: 20).
[41] Singer (2011: 50).

give preferential weight to human interests over like non-human interests – without thereby rejecting humanism. One could, for instance, think that human beings (simply *qua* human beings) have certain inviolable rights, without taking a stand on the (relative) moral standing of non-human animals. One could think that non-human animals have equal moral standing, though this would not entail that in all respects animals ought to be treated in the same way as human beings. Alternatively, one might even think that the idea of *relative* moral standing is incoherent.[42]

In fact, Singer does not think this: he advocates a "graduated view in which moral status depends on some aspects of cognitive ability, and that graduated view is applied both to humans and nonhumans."[43] Where the 'speciesist' takes moral standing to be indexed to species membership, Singer takes moral standing to be intimately related to the cluster of traits that he associates with 'personhood': self-consciousness, consciousness of oneself as a being persisting over time, and related features. Recall that Singer's principle said that the *like* interests of all those affected by a proposed action must be weighed equally. On Singer's view, not all interests are *like*; as we shall see, he thinks a 'merely' conscious being's interest in staying alive counts less than a self-conscious being's interest in staying alive. He draws the consequences regarding how relatively bad it is to kill a being of each category.

A further feature of Singer's theory is its more or less explicit commitment to the commensurability of all morally relevant phenomena. Singer's principle of equal

[42] See further the discussion of Korsgaard's views about moral standing in Chapter Two, Section IIB below.
[43] Singer (2009a: 575).

consideration is a measuring implement; it "acts like a pair of scales, weighing interests impartially. True scales favour the side where the interest is stronger or where several interests combine to outweigh a smaller number of similar interests, but they take no account of whose interests they are weighing."[44] The differences between interests are understood in quantitative terms. Even where interests are 'unlike', they are 'like' enough that 'smaller' interests might eventually outweigh larger ones. The idea that anything that is morally relevant could not *in principle* be measured, fed through the calculus, and maximised, is alien to Singer's moral theory. One of the central tenets of his utilitarianism, and thus of his form of anti-humanism, is his commitment to the basic calculability of the ethical domain. If some putatively ethical consideration cannot be forced into the mould of a preference or interest, if some putative wrong makes no difference to the "subjective experience" of the supposedly wronged party, it will not 'count' as ethically relevant after all.[45] In order to be calculable, all considerations must be reduced down to the common currency of utility.

In this respect, at least, Singer's views could not be further from those of Heidegger. This is not because Heidegger endorses an ethical theory opposed to Singer's, such as a form of deontology. On the contrary, as Sacha Golob notes, "Heidegger rarely engages directly with the familiar ethical or moral debates of the philosophical canon", in part because of his belief "that his own [ontological] concerns are explanatorily prior to such debates."[46] Nonetheless, it would be a mistake to conclude that Heidegger's work has no normative impetus. I shall have much more to say about Heidegger's 'ethics' in

[44] Singer (2011: 20-1).
[45] The quoted phrase comes from a personal communication from Singer to Jeremy Waldron, quoted in Waldron (2017: 242).
[46] Golob (2017: 623).

Chapter Two. For now, the most relevant point is that in the late 30s, Heidegger began to write in an increasingly polemical manner against what he called "calculation" (Ga65: §58-9; *QCT* 21, 23). His critique is closely related to the abiding concern displayed in his later work with the ascension of technology in modernity, and the concurrent "unbridled dominance of machination" (Ga65: 98, §59). Heidegger is not interested only in the proliferation of technological devices and machinery, their ever-increasing centrality to our daily lives. In fact, his primary interest lies in uncovering the relationship to Being that underlies this proliferation and the modes of thinking that accompany it. "The essence of technology," he tells us, "is by no means anything technological" (*QCT* 4). Instead, it is a "way of revealing" entities, according to which they show up to us increasingly only as "standing-reserve" [*Bestand*], resources or means to be ordered and manipulated: the "cultivation of the field" is transformed into a "mechanized food industry"; the Rhine river, the subject of a hymn by Hölderlin, is "dammed up into the power plant", or becomes "an object on call for inspection by a tour group ordered there by the vacation industry"; the wood in the forest becomes "cellulose, [...] challenged forth by the need for paper" (*QCT* 20, 15-18). Even the human being is eventually transformed into "raw material" for "consumption" (*OM* 86).

On Heidegger's view, these developments go hand in hand with a conception of the world as essentially *calculable*.[47] Heidegger has a complicated story to tell about the rise of this form of technological thinking and its relation to trends in modern metaphysical

[47] See Mulhall (2021), Essay Three, Part 2, especially Sections B and C for useful discussion of Heidegger's account of the age of technology and its "grounding projection of nature as inherently calculable" (2021: 264).

thought.[48] For Heidegger, philosophical responsiveness to these trends is of the utmost importance. "Meditation," he writes, "must strike up against the fact that the indifference with regard to Being" that characterises our age "is nothing less than the highest amplification of the power of *calculation*. The most indifferent and blindest denial of the incalculable is at work here" (Ga65: 351, §261). The understanding of the world as calculable, fully explicable by mathematical modes of thinking, conceals what Heidegger takes to be the proper subject matter of thought: Being, the 'incalculable'. Heidegger's complaint against humanism is of a piece with his concerns about technology: humanism will remain conditioned by metaphysics so long as its "determination of the essence of the human being [...] presupposes an interpretation of beings without asking about the truth of Being" (*LH* 245). Moreover, humanism's elevation of the human being and the technological conception of the world in terms of calculable and manipulable resources are two sides of the same coin. Humanism conceives of the human being as "the lord of beings"; Heidegger's aim is to reconceive of the human essence in terms of its relationship to Being. The human being is – must become – "the shepherd of Being" (*LH* 260).

While both exhibit concerns about the elevation of the human being, the distance between Singer and Heidegger's respective views should now be clear. Indeed, although Heidegger himself shows little interest in discussing utilitarianism as an ethical theory, his later work suggests that we could see its conception of what is morally relevant as symptomatic of the same larger historical trend in which he implicates humanism: part and parcel of the ascension of calculative rationality. If, as Heidegger claims, we find

[48] See Mulhall (2021: 265-7) for an account of the trajectory Heidegger traces in this respect from Descartes to Kant to Nietzsche.

ourselves entranced in modernity by a "bewitchery that directs everything toward calculation, utility, [...] manageability, and regulation" (Ga65: 98, §59), it will hardly be a surprise if our moral sensibilities also end up being refracted through the prism of technological thinking. Even if Heidegger did not have the moral sense of 'utility' in mind, its inclusion in his list is not inappropriate. Heidegger might perhaps have agreed – albeit for his own distinctive reasons – with Elizabeth Anscombe's judgement that consequentialism "is conceived perfectly in the spirit of the time and might be called the philosophy of the flattery of that spirit."[49]

Singer and Heidegger's respective forms of anti-humanism could not be further apart. Nonetheless, it is not clear that either position represents an unequivocal advance in regard to the two central problems facing humanism.[50] Singer's view is motivated in part by an admirable opposition to anthropocentrism. However, as I noted above, he remains committed to the idea that the superior cognitive capacities associated with 'personhood' – self-consciousness, perception of oneself as a being persisting in time, advanced linguistic capacities etc. – grant a higher moral status to the beings that possess them. Singer draws consequences from this commitment that many of his critics – both philosophers and activists – have found abhorrent.[51] In particular, Singer advocates a highly controversial position regarding the wrongness of killing, according to which "the lives of some animals, with higher cognitive capacities, should be regarded more seriously than we do at present, while taking the lives of some human beings,

[49] Anscombe (2005: 168).
[50] My critical appraisal of both thinkers in this section will be unfortunately but unavoidably brief. See Chapter Two for a more detailed, textually supported discussion of Heidegger's critique of human rights and the relationship of his work to 'the inhuman'. I hope to elaborate on the criticisms of Singer advanced here in future work.
[51] See e.g. Johnson (2009); Kittay (2009); Crary (2019b); Waldron (2017), Chapter 6.

lacking in such capacities, should be seen more permissively."[52] On these grounds, Singer goes so far as to endorse the infanticide of babies born with very severe cognitive disabilities – a position that has led critics to accuse him of ableism and even eugenicism.[53]

In the light of this 'more permissive' attitude towards taking the lives of human beings with disabilities, it is very difficult to take seriously Singer's claim that "the aim of [his] argument is to elevate the status of animals rather than to lower the status of any humans."[54] His theory institutes precisely the kind of hierarchy of moral status among human beings with which influential forms of 'humanism' have been charged. Moreover, while Singer's account clearly takes decisive steps towards a less anthropocentric ethics, it is not clear whether he goes far enough. The quotation above suggests that, when the question is one of taking life, it is the animals with sufficiently developed capacities – the Great Apes, dolphins, perhaps pigs – that benefit most from the elevation of moral status that Singer proposes. While he does also denounce the cruelty of factory farming, he does not seem to think there is anything intrinsically wrong with killing an animal that is not self-conscious: "For a *merely* conscious being, death is the cessation of experiences. [...] Death cannot be contrary to an interest in continued life any more than birth could be in accordance with an interest in commencing life."[55] Singer suggests that it would not be inappropriate to view such creatures as "receptacles for experiences of pleasure and pain" – mere vessels of utility,

[52] Singer (2009b: 34).
[53] See Singer (2009b: 47-51) for Singer's own (unsurprisingly partisan) account of some of the protests against his work.
[54] Singer (2011: 67).
[55] Singer (2011: 112, emphasis added).

of abstracted moral 'value'.[56] Those who are inclined to see animals as inherently morally significant – as exerting a moral claim on us just in virtue of the fact that they are living, vulnerable creatures with their own ends and their own 'good' – might have hoped for more.[57]

What of Heidegger? Heidegger is concerned to avoid "elevating the human being to the center of beings", and to that extent his critique of humanism seems to set itself in opposition to anthropocentrism (*LH* 268). However, his reason for that opposition has little, if anything, to do with the moral claims of non-human animals. His primary concern is for human beings to avoid persisting in the "homelessness" that is a "symptom of oblivion of Being", and instead to return to the "homeland" of the "nearness to Being" (*LH* 258). Human beings can come into this proximity to Being because they are the only beings who have an understanding of Being, and hence a 'world' in Heidegger's special sense. This is a privilege that Heidegger explicitly withholds from non-human animals, writing that other "living creatures" are "separated from our ek-sistent essence by an abyss" (*LH* 248). For this reason, Heidegger's critique of humanism has been accused of retaining a highly anthropocentric outlook, albeit one based on a transformed conception of the human essence.[58] Furthermore, Heidegger's association with National Socialism in the early 30s, coupled with the anti-Semitism explicitly expressed in some of his posthumous writings and his persistent

[56] Singer (2011: 112); compare his 'bottle' metaphor in Coetzee (1999: 90). See also Mulhall, (2008b: 65-7) for critical reflections on the latter passage.
[57] See Korsgaard (2018), Chapter 9, section 9.2.3, for criticism of Singer's 'receptacle' view; see further Chapter 2 for her account of what it is for an animal to have a 'good'. See also Donaldson and Kymlicka (2011: 260-1, fn. 8, 263, fn. 1) for some brief critical remarks regarding Singer's position from an Animal Rights perspective, and Chapter 2 for their defence of universal basic rights for animals.
[58] See e.g. Derrida (1969: 50-1); (1989), Chapter VI; (2008) *passim*; Glendinning (1998), Chapter 4. See also the citations in Chapter Two, fn. 3 below.

centring of the German people, should – at the very least – warrant serious concern about whether his rejection of humanism can properly be seen as an advance on it.[59]

Heidegger soon came to reject National Socialism as a viable alternative to modernity – indeed, he came to view it as one more devastating consequence of the present technological epoch of the history of Being. But the very character of his later critique threatens to minimize the responsibility of those who perpetrated the holocaust. The latter becomes just one more effect of the depersonalised global domination of technology: "Agriculture is now a mechanized food industry, in essence the same as the production of corpses in the gas chambers and extermination camps, the same as the blockading and starving of countries, the same as the production of hydrogen bombs" (Ga79: 27).[60] Now, 'in essence the same' does not mean 'equal in moral significance' for Heidegger; it means these events all stem from the same essential process (namely, the essence of technology). But this does not amount to much in the way of consolation or defence, especially because Heidegger seems to abjure the kind of normative resources that might allow us to make judgements of relative moral significance. He downplays the differences between fascist and democratic forms of social organization, suggesting that "the devastation [of the modern age] reigns also and precisely [...] where human rights are respected, where civil order is maintained" (Ga77: 139).[61] Finally,

[59] See Golob (2017: 624, 633) for brief but trenchant comments on the anti-Semitism evident in Heidegger's posthumously published *Schwarze Hefte*; see also Wolin, ed. (1993) for an earlier collection of primary texts and critical essays bearing on Heidegger's relation to National Socialism. See further Chapter Two, Section III below.

[60] This sentence is from an earlier version of Heidegger's essay 'The Question Concerning Technology', first delivered as a lecture in Bremen in 1949. The claim about the gas chambers is not repeated in the later version of the text, and is one of only a handful of mentions of the holocaust in Heidegger's *Gesamtausgabe*.

[61] See also Heidegger's comments on democracy and other political systems in Ga69: 160-2; OAG 104.

although in the 'Letter on "Humanism"' he repeatedly distances himself from advocacy of 'the inhuman' (*LH* 244, 251, 265), in a posthumously published treatise he seems to suggest just the opposite, identifying the kind of thinking that he is calling for as "in-human" and giving the following gloss on that term: "does not turn to standards or goals or motivations belonging to humankind hitherto" (Ga69: 23).[62] The result is what appears to be an amoralism that deprives us of the normative resources that might provide effective opposition to 'the inhuman'.[63]

In this section I have presented two very different anti-humanist viewpoints, both of which nonetheless appear to be vulnerable to concerns that are closely analogous to those facing humanism. Can we draw any more general inferences from this discussion regarding the prospects for anti-humanism? We would not be warranted in concluding that *no* form of anti-humanism could improve on the failures of existing forms of humanism. Nor can we conclude that these ethically questionable implications are a necessary feature of anti-humanism. Nonetheless, in neither Singer's nor Heidegger's work do they seem entirely incidental. In each case, though in different ways, there appears to be a withdrawal of resources (such as the notion of human rights) that might

[62] In his review of this translation, Mark Wrathall suggests that the gloss cited in the main text pertains not to Heidegger's own ideal of thinking, but to our experience of "ourselves, our purpose in life, and the entities around us" when we find ourselves in a world organized by the imperative of the will to power; see Wrathall (2016). Wrathall thus views the retraction of recognizable human standards as a symptom of the modern age, as diagnosed by Heidegger. In context, however, this cannot be the right reading, not least because Heidegger immediately goes on to identify such 'in-human' thinking with *Da-sein*, a term which by this point in Heidegger's intellectual development has come to mean a mode of being that is the "highest one of humans, namely the possibility of grounding and preserving truth itself" (Ga65: 237, §176); see Chapter Two, Sections I and IIA below for discussion of this shift in Heidegger's terminology. In other words, it is precisely the thinking that preserves the truth of Being that Heidegger is here suggesting cannot be made answerable to the standards held by human beings 'hitherto'.

[63] One of my central concerns in Chapter Two is to determine the extent to which this appearance is *mere* appearance.

in principle serve as a bulwark against various forms of inhumanity.[64] To anyone receptive to the charges levelled against humanism, it will be difficult to represent this withdrawal as anything other than a loss or an impoverishment of our moral resources, for reasons that I shall try to articulate in the next section. For this reason, our preference in the first instance should be to see how far we can go in revising humanism and making it responsive to the charges of inegalitarian hierarchism and anthropocentrism, rather than prematurely dismissing it *tout court*. For now, we can draw the more modest conclusion that the onus is on the anti-humanist to show that their position does not have inegalitarian, anthropocentric, or inhuman consequences. If they cannot do so, their view will not represent genuine progress over humanism, and may even amount to a form of regression.

IV. Humanism as a Basic Commitment?

One way to bring out the sense that the rejection of 'the human' as a normatively significant concept would amount to an impoverishment of our moral resources would be to consider the roles it currently plays in our ethical and political projects – including in our critical practices. For there is surely a difference between abandoning or rejecting the normative uses of a concept in contexts where it is already playing an important role, and wondering whether – in another world – we could have done without it (or even could have done *better* without it). We cannot simply select our concepts *ex nihilo*. It therefore cannot be a matter of indifference that appeals to the notion of the human are pervasively embedded in our linguistic and legal practices.

[64] Compare here Adrian Moore's discussion of Spinoza's 'anthropocentrism', which Moore argues acts as a bulwark against what would otherwise be a sinister relativism (2017: 51-2).

The demand that one be 'treated as a human being' is a common way of registering and responding to certain forms of injustice and maltreatment, especially those that one might view as paradigmatically dehumanising. The experience of dehumanisation recounted in the work of Jewish writers who survived persecution in Nazi Germany is a case in point. Primo Levi describes an encounter with a supervisor who is looking over Levi's credentials, and who has the power to move him out of the concentration camp so that he can work in a chemical factory instead. Levi writes that the look that passes between him and this supervisor "was not one between two men."[65] As Cora Diamond puts it, Levi is confronted by "a look that leaves [him] a mere circumstance and not a human being."[66] Emmanuel Levinas gives a comparable account of his experience as a prisoner of war in Nazi Germany: "the other men, called free, who had dealings with us or gave us work or orders or even a smile – and the children and women who passed by and sometimes raised their eyes – stripped us of our human skin."[67] These specific encounters are not overtly violent, perhaps at times even involve an appearance of cordiality ('even a smile') – and yet they are radically degrading, profoundly lacking in the respect that we take (or ought to take) to be due to all human beings. It is not clear whether the same experience could be articulated without the concept of the human.[68]

The worry about anti-humanism is that it will leave us unable to adequately register or do justice to these basic moral considerations. That does not mean that all is well with the many and varied uses to which the concept of the human is put, or that we should

[65] Levi (1979: 111).
[66] Diamond (1988: 265).
[67] Levinas (1990: 153).
[68] This is related to the fact that, as Alice Crary has put it, "Nazi methods were designed to target specifically human susceptibilities [and] Nazi instruments of terror served to destroy distinctively human forms of dignity that essentially involve both social recognition and self-respect" (2019a: 240-1).

not subject those uses to careful criticism. But it does tell us something about the kind of criticism for which we should be striving. Versions of the first charge against extant forms of humanism – that it has instituted hierarchies among human beings – are serious, but as I suggested in Section II, they are also best understood as internal criticisms. For the point of the criticism is often that the operative notion of the human has been too narrow – has failed to acknowledge the *humanity* of some groups of human beings.

The second charge – that of anthropocentrism – must also be taken seriously. But it should be taken as an invitation to expand the sphere of moral concern: to elevate the status of the non-human, not to diminish the status of the human. As we saw above, while Singer claims only to be doing the former, his claim is belied by the judgments he actually endorses concerning human beings with cognitive disabilities. But morality is not a zero sum game. On the one hand, our acknowledgement of the claims of non-human animals should not be traded off against the claims of any human beings, especially those who already occupy a highly precarious social position. On the other hand, recognising human rights need not mean denying 'animal rights'. Finally, accepting a discourse and legislation of animal rights – if this turns out to be the best way of codifying our collective responsibilities and duties to animals – need not mean recasting human rights in the terms of a more universal legislature. It may be that we owe some things to our fellow humans that we do not owe to our fellow animals, both because some human beings are involved in forms of life in which no non-human is involved in a comparable way (such as religion), and because human beings may be

susceptible to specifically human forms of degradation and violence. Even so, none of this detracts from the idea that non-human animals exert a real ethical claim on us.

The concept of the human seems to have a special status. This comes out in a passage from Bernard Williams, in which he is discussing those regions of our ethical outlook where a 'crisis of legitimation' is unlikely to arise:

> We believe [...] that in some sense every citizen, indeed every human being – some people, more extravagantly, would say every sentient being – deserves equal consideration. Perhaps this is less a propositional belief than the schema of various arguments. But in either case it can seem, at least in its most central and unspecific form, *unhintergehbar*: there is nothing more basic in terms of which to justify it. We know that most people in the past have not shared it; we know that there are others in the world who do not share it now. But for us, it is simply there.[69]

For Williams, then, some elements of our "ethical and political outlook" will appear to us as non-negotiable.[70] Williams expresses some hesitancy as to how best to characterise those elements: are they better understood as propositional beliefs, or as 'schemata' of arguments (whatever this latter notion might come to)? To remain neutral on this question and others in the vicinity, I shall call such elements 'basic commitments'. Whether we construe them as beliefs or commitments of some other kind, they represent the point at which we hit bedrock and our spade turns, ethically speaking.[71]

---

[69] Williams (2006: 194-5).
[70] Williams (2006: 196).
[71] See Wittgenstein, *PI* §217 for the origin of the 'bedrock' metaphor; see the Bibliography for a list of abbreviations of works by Ludwig Wittgenstien. It is no coincidence that Williams' phrase 'it is simply there' strongly recalls what Wittgenstein is inclined to say when his spade turns: "This is simply what I do." On the contrary, shortly after the passage cited in the main text, Williams acknowledges his debt to Wittgenstein, who "correctly insisted that there was an end to justifications" (2006: 196). Nonetheless, here as elsewhere, Williams distances himself from what he takes to be Wittgenstein's excessively wide understanding of the first person plural and a corresponding "conception of philosophy's subject matter as exclusively *a priori*"

In some sense we recognise alternatives to such notions, since we know that not everyone shares our commitment. However, in Williams' view those alternatives have a very specific status. We cannot understand them as live options – genuine ethical alternatives *for us*. There is an asymmetry between the kind of sense our basic commitments make to us, and the kind of sense we can make of the alternatives to those commitments. Williams suggests that while we can recognise the contingency of our commitments, can indeed imagine a future in which we no longer accept them, those commitments themselves place constraints on the terms in which we could envisage such a situation:

> We cannot in our thought go beyond our outlook into the future and remain identified with the result: that is to say, we cannot overcome our outlook. If a possible future that figures in those shadowy speculations does not embody some interpretation of these central elements of our outlook, then it may make empirical sense to us – we can see how someone could get there – but it makes no ethical sense to us, except as a scene of retrogression, or desolation, or loss.[72]

Williams' example of a basic commitment here is the notion of equal consideration. However, we can also fruitfully understand our commitment to the claim that 'human being' is a normative concept – that is, our commitment to humanism – in terms of his line of thought. I would suggest that the ethical concept of 'the human' is a part of 'our' outlook in just the way Williams describes. A situation in which 'we' collectively no longer take this concept to have an ethical meaning is not unimaginable. We can imagine empirical or ideological processes by which such a situation could come about:

(2006: 196); see also Williams (1981). Williams' 'we' is self-consciously "local" or historically situated; furthermore, he is concerned in the present context to defend the relevance of history – and thus of *a posteriori* considerations – to the philosophical project of making sense of ourselves. While this last point strikes me as highly plausible, it is worth noting that there are much more attractive readings of Wittgenstein's 'we' available; for that reason, Williams' criticisms are arguably misguided. See Chapter Four, Section I below for further discussion.

[72] Williams (2006: 197).

propaganda would be an obvious example of a mechanism or strategy by which our commitment to this concept could be – in various times and places, has been – eroded. But to imagine a process by which we – many of us – succumb to such propaganda is of necessity to imagine a process with whose outcome we cannot, ethically speaking, identify.[73] It is to imagine that we have been misled or duped or manipulated without our knowing it. It is to imagine 'a scene of retrogression, or desolation, or loss' – or perhaps distortion, or corruption, or perversion.

It may be objected that there is a worrying conservatism lurking in the background of the notion of a basic commitment to humanism sketched here, or at least a failure to properly acknowledge the severity of the charges facing humanism. (This worry will perhaps only be intensified if we are tempted to construe Williams' commitment to egalitarian humanism as inextricable from his commitment to liberalism as a political outlook.) To continue the Wittgensteinian metaphor, it is certainly open to us to ask: "But does our spade turn too soon here?"[74] This is a possibility we should take seriously. Nonetheless, there are two considerations that should be borne in mind, and that may go some way toward allaying the worry about conservatism.

The first point is that genuine commitment to humanism means commitment to progressive action: to the transformation of society, the elimination of oppression and domination, the realization of a more egalitarian condition. Indeed, this is one reason for viewing the possible loss of our humanistic commitments as a scene of regression. For as Rainer Forst has recently argued, we should view regression as having an

[73] Compare here Diamond (1998: 260).
[74] I borrow this formulation from Jaeggi (2018: 20).

important futural aspect. To regard a transformation as regressive is to regard it as involving not just the reversal of progress that has already been achieved, but also the loss of as-yet-unrealised standards oriented toward the future.[75] Humanistic ideals are not undemanding. Even where everyone pays lip service to the idea that all human beings matter, it in no way follows that humanistic ideals have been achieved. Even where such ideals are enshrined in law, it in no way follows that those laws are respected; they are violated every day by representatives of the state whose profession is to enforce them. Humanistic ideals may never be fully realised. But part of what is being 'conserved' in the commitment to humanism is a commitment to progress – a commitment to improving the condition of all human beings. In this sense, the loss of our basic commitment to humanism would be a loss not just of the progress that has been achieved, but of progress to come – the loss of a possible future, fragile and precarious though that possibility may be. We know from bitter experience what regression is and what inhumanity is.[76] This is one way in which our knowledge of our history, of 'how we got here', is centrally relevant to our ethical self-understanding. Normative theorising can and should begin with that knowledge.

The second point is that Williams is not claiming that our basic commitments cannot be subject to any critical revision. Williams' claim that any future with which we could identify must 'embody *some* interpretation' of our basic commitments suggests that there is a measure of flexibility to our self-understanding. Indeed, a new interpretation of our commitments might not just be acceptable to us – it could conceivably appear as

[75] Rainer Forst, 'Regression, Reason and Democracy', paper delivered at the "Philosophy and Social Science" Conference, University College Dublin, Saturday 19th May 2022.
[76] See Adorno (2000: 175); see also Freyenhagen (2013), Chapter 5.

a natural or even a necessary extension of our present outlook. It is surely significant that Williams interpolates a line of dissent into his very formulation of 'our' basic commitment to equal consideration – does it extend just to human beings, or to all sentient beings? (If the latter, would we have to understand that commitment along the lines suggested by Singer, or are other interpretations of 'equal consideration' available and preferable?) Returning to 'the human', the point is not that our understanding of that notion's ethical content is immune from revision; the point is that although 'our' outlook could be transformed in ways with which we, now, could still conceivably identify, it would not survive the obliteration of our sense *that* the concept of the human is of basic ethical significance. To view humanism as a basic commitment means demanding that concern for all human beings remain integrated into any future perspective we could recognise as 'ours'. It means viewing it as a constraint on the ethical plausibility of any moral or philosophical theory that it 'embody some interpretation' – some ethical interpretation – of the concept of humanity.

The pressing question thus becomes: What interpretation of this concept should we accept? How should we make sense of our humanistic commitments? Our theoretical or philosophical claims about human beings, whether these are claims about what human beings essentially are, or about why they matter, or both, are likely to inform and inflect our beliefs about what human beings (should) want or need, our understanding of what it is to live in community with other human beings, and even our judgements about who properly 'counts' as a human being. Adopting a distorted philosophical conception of the human is thus likely to have wide-ranging ramifications for our

attitude towards and treatment of one another. Getting our ethical concept of the human being clearly into view is thus an important philosophical task.

That task will involve, in part, coming to see the relationship between humanity and animality in the right light. For one of the obvious facts to which a philosophical account of the human must do justice is the fact that human beings are themselves animals. Any form of humanism – or, for that matter, anti-humanism – that neglects, suppresses, or distorts this fact is likely to go astray.

The challenge, then, will be to articulate a conception of 'the human' that successfully records and renews our sense that human beings matter, without denigrating other animals or inadvertently instituting hierarchies among human beings. Bearing the charges set out in Section II in mind, in the following chapters I will explore the normative conceptions of the human found in the work of four thinkers: Immanuel Kant, Martin Heidegger, Judith Butler, and Ludwig Wittgenstein.

In Chapter One, I argue that Kant's appeal to rationality as the ground of human beings' moral standing makes it difficult for a distinctively Kantian form of humanism to avoid the charges introduced in Section II. In Chapter Two, I consider Heidegger's distinctive critique of humanism, evaluating his claim that he nonetheless does not advocate the inhuman. I argue that while Heidegger can defend this claim if we adopt the ontological understanding of 'inhumanity' that his later work seeks to articulate, he nonetheless appears to deprive us of the normative resources that might allow us to resist inhumanity in the ordinary moral sense. In Chapter Three, I turn to Butler's

recent work on the ontology of the body. I argue that their ontological claims about corporeal vulnerability, interdependency, and precariousness provide us with resources for a more critical and egalitarian form of humanism, despite their own ambivalence concerning humanism generally. Insofar as the ontological condition that Butler identifies is shared by non-human animals, our acknowledgement of it also provides a potential source of solidarity with other embodied creatures, thus allowing their account to avoid the anthropocentric implications of other forms of humanism. Finally, in Chapter Four I use work by Wittgenstein and Stanley Cavell to provide an account of the philosophical 'we' and the universal voice of philosophical claims. I argue that the comparison Cavell makes between aesthetic and philosophical claims can shed light on the aims and methodology of Butler's ontology, while also providing an attractive way of understanding the project of giving a philosophical account of ourselves more generally. Finally, I conclude by suggesting that although Wittgenstein's work is sometimes read as problematically anthropocentric, it can nonetheless provide support for the project of developing a less anthropocentric humanism. It does so by raising questions about interpersonal understanding and our understanding of other animals; by drawing our attention to the affinities and resemblances between human and non-human forms of life; and by offering us a picture of mental concepts that gives a central place to the body and its natural expressiveness.

Chapter One
Kant's Humanism and Kantian Humanism: Humanity, Rationality, and Ends in Themselves

One of the most celebrated doctrines of Immanuel Kant's moral philosophy is his notion, set out in the 'formula of humanity', that human beings must always be treated as ends in themselves, never merely as means.[1] As Lucy Allais has recently put it, Kant is "a theorist of an incredibly influential (arguably, revolutionary) moral philosophy centered essentially around the idea of human freedom and equality which argues inflexibly that it is always wrong to treat a person as a means and to fail to respect their humanity."[2] This notion of respect for humanity and human autonomy has been employed in ways that exhibit its radical potential, and that go beyond what Kant himself envisioned.[3] Some philosophers have even seen in Kant's work the promise of an underpinning for basic human equality, dignity, and rights. In Jeremy Waldron's words, "If we are pursuing distinctive human equality, then [Kant's theory] is the sort of account we are looking for."[4] Nonetheless, there are reasons to be cautious about the prospects for Kant's defence of the moral importance of humanity. In the previous chapter, I set out two worries about historically influential forms of humanism: that they have been unacceptably *hierarchical*, failing to give equal moral standing to all

[1] See Immanuel Kant, *Groundwork of the Metaphysics of Morals* (1785), 4: 429 for his canonical formulation. With the exception of citations of the *Critique of Pure Reason* (1781/7), which I give in the standard A/B format corresponding to the pagination of the first and second editions, all references to Kant's works give the volume and page numbers of the *Akademie* edition of his oeuvre, which are supplied in the margins of most English translations. Unless otherwise noted, all translations are from *The Cambridge Edition of the Works of Immanuel Kant*.
[2] Allais (2016: 1).
[3] See e.g. Nussbaum (1995); Langton (2009b); Forst (2017).
[4] Waldron (2017: 101). Recall that for Waldron, 'distinctive' suggests a ground of equality that raises us above other animals; see Introduction, fn. 33 above. This will be relevant later on.

human beings, and that they have been unacceptably *anthropocentric*, failing to give appropriate moral standing to non-human animals. Kant's version of humanism has been the target of both charges, and there are identifiably hierarchical and anthropocentric strands in his ethical and political writings.

Firstly, there are reasons to think that Kant's own humanism was not as egalitarian as it might appear at first glance. Kant's apparently universalistic ethical doctrine coexisted with a hierarchical view of race. Furthermore, despite the fact that slavery is surely one of the most glaring ways in which a person might be used merely as a means, it is both notable and disturbing that Kant failed – with the exception of some brief comments in his very late writings – to condemn the transatlantic slave trade.[5] Both Kant's critics and those broadly sympathetic to his work – Allais is among the latter – have sought in recent years to make sense of this apparently blatant contradiction in Kant's thought, between his racism on the one hand and his official egalitarianism on the other. Similar discussions have centred on Kant's sexism. In both the *Metaphysics of Morals* and his *Anthropology from a Pragmatic Point of View* (1798) he mentions "all women" in the same breath as minors in his discussion of "passive citizens" who "lack civil personality" and therefore should not have the right to vote, on the grounds of their alleged "immaturity" (6:314; 7:208-9).[6] Finally, there are passages in Kant's work that

[5] This is in spite of the fact that in the *Metaphysics of Morals* (1797), Kant explicitly advanced an argument to the effect that no one could genuinely consent to being a slave by renouncing their own freedom: "A contract by which one party would completely renounce its freedom for the other's advantage would be self-contradictory, that is, null and void, since by it one party would cease to be a person" (6:283). One would think that this provides an *a fortiori* argument for the wrong of slavery in general: if slavery would be wrong even in the supposedly consensual or contractual case, it could hardly be permissible in the case of the overt expropriation of a person's freedom.

[6] Pauline Kleingeld notes that although there is evidence that Kant "dropped the thesis of racial hierarchy and began to criticize European colonialism" in later life, "he never made parallel revisions to his account of the status of women"; see Kleingeld (2019: 3). Kleingeld's own aim is

suggest that he attributed inferior moral standing to human beings with certain kinds of 'mental derangement' (Kant's term). This final exclusionary gesture has been less frequently discussed, but I shall argue that it sheds light on the limitations of Kant's version of humanism.[7] In particular, it highlights the fact that Kant takes our moral standing – our status as ends in ourselves – to be grounded in our *rational* nature.

This account of moral standing also underpins the second exclusionary aspect of Kant's theory: his notoriously anthropocentric account of our duties to non-human animals. On Kant's account, all such duties are indirect – they are not really owed to the animals as such, but are instead duties that we owe to ourselves. As Kant puts the point in a striking passage from his *Lectures on Ethics*:

> We have duties to animals, in that we thereby promote the cause of humanity. So if a man has his dog shot, because it can no longer earn a living for him, he is by no means in breach of any duty to the dog, since the latter is incapable of judgement, but he thereby damages the kindly and humane qualities in himself, which he ought to exercise in virtue of his duties to mankind. (27:459)

Three features of this passage are worth noting right away. First, while Kant does claim that we have duties *regarding* the treatment of non-human animals, he also makes the counterintuitive claim that they are not strictly speaking duties owed *to* those animals. I do not owe it to my dog not to shoot 'it'; I owe it, if at all, to myself and to 'mankind' in general. Second, Kant offers a brief explanation as to *why* I owe nothing to my dog: it is because the dog is 'incapable of judgement'. In Kant's view, obligation bears an intimate relation to the capacity for judgement, and indeed to 'rational nature' in general. (I shall explain Kant's reasons for this view in the next section.) Finally, Kant

---

to determine the extent to which we can "separate the principles from the objectionable views and use Kant's principles to criticize his own biases" (2019: 4).

[7] See Frierson (2014), Chapter 6, especially 6.2-6.5, for a discussion of Kant's views on 'mental disorder'. However, Frierson is more concerned with the bearing of this empirical material on Kant's theoretical philosophy than with its ethical implications.

offers an explanation for why we owe it to ourselves not to mistreat animals: it is because it damages our 'kindly and humane' qualities, which we owe it to each other to develop.

This final aspect of Kant's view has recently been criticised – indeed, charged with a kind of incoherence – by Christine Korsgaard, one of the most prominent figures in contemporary Kantian ethics, in her book *Fellow Creatures*.[8] While sharply criticising Kant's own explanation of our obligations to animals, Korsgaard nonetheless seeks to provide a superior account of those obligations – one that seeks to explain, from what Korsgaard takes to be broadly Kantian premises, why we owe them directly to the animals themselves. Her argument turns on the idea that we are obliged to value our own animal nature – to value ourselves as ends *qua* animals. Korsgaard thinks we can and must generalise this idea, and come to regard other animals as ends in themselves too.[9] Korsgaard's project in *Fellow Creatures* provides an especially stark example of the sometimes vast gulf between Kant's ethics and Kantian ethics.[10] Applying this familiar distinction, we might fruitfully draw a contrast between Kant's humanism and Kantian humanism. If the charges briefly canvassed above are just, Kant's own humanism is morally implausible. But this leaves open the question of whether we can formulate a

[8] See Korsgaard (2018), Chapter 6. I discuss her critique in Section II.

[9] Korsgaard (2018), especially Chapter 8. See Section II for a critical discussion of Korsgaard's argument.

[10] See Allais (2016: 31ff.) for a brief discussion of this distinction. Allais points out that the term 'Kantian ethics' can itself be understood in at least two ways: it "is sometimes taken to mean something broadly inspired by Kant [but] not strictly concerned with all of his theoretical commitments, but it could also mean a theoretical system proposed by Kant which, when consistently worked out, has results at odds with some of Kant's own actual thoughts" (2016: 31). However, separating what is properly speaking a theoretical commitment out from what is a first order ethical judgement (an 'actual thought') is not necessarily a straightforward matter. (Is the idea that I owe no duties to my dog, because 'it' is 'incapable of judgement', an example of the former or the latter?) Thus, the line between the two kinds of project that might be called 'Kantian ethics' is perhaps not especially sharp.

recognisably Kantian humanism that is ethically defensible – one that does justice to the idea that every human being has unconditional moral standing, without excluding non-human animals from the sphere of moral concern.

We could perhaps view Korsgaard's own project in this light, though one might have an initial reservation concerning whether her position is aptly called a form of humanism at all. The thrust of her argument is to show that both humans and non-human animals have moral standing, and that we cannot say that humans are absolutely more important than animals. Is that not simply a denial of 'humanism'? According to the inclusive definition of humanism mooted in Section I of the Introduction, it is not. Humanism, I argued, is best understood not as the view that human beings are *uniquely* important, but that 'human being' is a normatively relevant notion: all human beings are inherently important, *matter* morally speaking, in virtue of being human beings. Different 'humanisms' will return different answers to the question of what, if anything, grounds or explains or justifies this importance, the moral relevance of humanity. Some such answers will be incompatible with the idea that non-human animals are also inherently morally important 'ends in themselves', or at any rate that they are equally as important as humans, but other answers will seek to make that idea comprehensible. Kant's answer is of the former kind; Korsgaard's is of the latter. But insofar as it remains committed to the moral standing of all human beings – and indeed provides an account of why some of our obligations to human beings differ markedly from our obligations to other animals – Korsgaard's position remains recognisably humanistic.

I am now in a position to articulate the project of this chapter, which is to determine the prospects for a recognisably Kantian form of humanism. To what extent can such a position answer the charges of inegalitarianism and anthropocentrism levelled against Kant's own account? The chapter is divided into two sections, addressing each of those charges respectively. I adopt a similar strategy in each part: first, I set out what I take to be the exclusionary claims made by Kant himself; second, I identify the deeper tendencies in Kant's thought that underlie those claims, and draw out some tensions in his view; third, I examine some contemporary Kantian positions that in various ways improve on Kant's theory in relation to these issues; and finally, I explain some reservations one might have about such strategies. In neither part do I offer a 'knockdown' argument against Kantian humanism. Nonetheless, I argue that the considerations I bring to bear against Kantian versions of humanism cast serious doubt on the adequacy of Kantian premises in making sense of our obligations to certain human beings and to non-human animals. To the extent that views like Korsgaard's account for the latter, they do so by departing to a significant degree from Kant's fundamental theoretical commitments – such that we might wonder whether such accounts really remain 'Kantian' after all.

I. Hierarchical Strands in Kant's Humanism?

Kant draws two distinctions in the *Groundwork* that provide a crucial underpinning for his claims about the moral status of both human beings and non-human animals. The first is between *persons* and *things*, the second between *dignity* and *price*. These two distinctions turn out to be correlated with each other, via Kant's notion of an end in itself:

> Beings the existence of which rests not on our will but on nature, if they are beings without reason, still have only a relative worth, as means, and are therefore called *things*, whereas rational beings are called *persons* because their nature already marks them out as an end in itself, that is, as something that may not be used merely as a means, and hence so far limits all choice (and is an object of respect). (4:428)

> In the kingdom of ends everything has either a *price* or a *dignity*. What has a price can be replaced with something else, as its *equivalent*; what on the other hand is raised above all price and therefore admits of no equivalent has a dignity. [...] That which constitutes the condition under which alone something can be an end in itself has not merely a relative worth, that is, a price, but an inner worth, that is, *dignity*. (4:434-5)

What is the condition under which alone a being can be an end in itself, the condition under which a being can have dignity? Here is Kant's answer:

> Morality is the condition under which alone a rational being can be an end in itself, since only through this is it possible to be a lawgiving member in the kingdom of ends. Hence morality, and humanity *insofar as it is capable of morality*, is that which alone has dignity. (4:435, emphasis added)

I shall return to the notion of the kingdom of ends towards the end of this chapter. For now, I want to draw attention to the relationship Kant indicates between humanity and the various other moral concepts in play. Kant thinks that it is the capacity for morality that confers dignity on human beings. Since human beings have dignity, they are 'raised above all price'; they have an inner rather than a merely relative worth, and therefore cannot be treated as mere means. On Kant's conception, our dignity – our standing as ends in ourselves – is inextricable from our status as *persons*, rational beings. For to be capable of morality is to be an autonomous 'lawgiver': to be able to make the moral law the basis for one's action. Practical reason – indeed, *pure* practical reason – is a prerequisite of morality. Human beings have dignity, are ends in themselves, insofar as they are moral beings; they are moral beings insofar as they are *rational* beings.

Thus, it is no surprise that Kant refers to the notion of a *person* in the canonical statement of his 'formula of humanity': "*So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means*" (4:429, emphasis in original). In the *Groundwork*, then, humanity is associated closely with rationality, and Kant shifts easily from talking about 'humanity' (*Menschheit*) to the more general idea of a person (*Person*). Indeed, it is unclear whether '*Menschheit*' refers to the idea of the humanity of a human being, 'Man' (*Mensch*) – roughly understood as a biological determination – or to the personhood of any rational being (*Person*).[11] Indeed, Kant takes pains to emphasise that the categorical imperative applies equally to non-human persons, should we ever encounter them (see e.g. 4: 408; 411-2; 425; 428-31).[12]

Kant describes his formulation as the "principle of humanity, and in general of every rational nature, *as an end in itself*" (4:431). His explanation for the *a priori* status of this principle – the fact that it is not "borrowed from experience" – makes an appeal to its universality: "it applies to all rational beings as such and no experience is sufficient to determine anything about them" (4:431). Clearly, the principle is supposed to have a very general application, containing all rational beings within its scope. Indeed, this

[11] On this point I would part ways with Charles Mills, who cashes out the distinction as follows: "'Persons' is the non-sexist way of referring to humans instead of calling them 'men'"; see Mills (2017: 92). On my view, the non-sexist rendering is 'human beings' (or just 'humans'), while 'person' denotes a concept that does not completely overlap with our concept of a human being. Persons are autonomous rational beings, some of which could in principle be non-human; human beings are generally autonomous rational beings, persons, but it is not clear whether they always are. This, I think, is Kant's view. As we shall see, it is also very close to what Mills takes to be Kant's view. In any case, it is worth drawing the distinction clearly, because it has consequences for what we might think is salvageable or worth reconstructing from Kant's account.

[12] See Derrida (1969: 42-3, fn. 12) for comments on these passages. Derrida suggests that for Kant, the human being remains the central point of reference even when he appears to point to rational beings 'in general': "in Kant's discourse, 'reasonable being' and 'humanity' are always associated by the conjunction 'and'" (1969: 43, fn. 12).

general application seems to be the explanation for its application to human beings: it applies to human beings insofar as they are also rational beings. So far, Kant's position may appear to be a general and egalitarian form of humanism. Human beings must be regarded and treated as ends in themselves because their humanity must be respected; their humanity must be respected because rational nature in general must be respected, and we find rational nature in human beings. However, what is not yet clear is whether Kant has or recognises grounds for thinking that *all* human beings are rational beings, persons.

A. Are All Human Beings Rational Beings?

If it is an open question whether a particular human being is a rational being, then it will also be an open question whether all human beings fall under the scope of the formula of humanity. The question of whether Kant viewed all human beings as rational beings, as persons, is thus a pressing one. However, it is not immediately clear what, on Kant's view, the criteria for rationality might be. One might expect such criteria to be empirical, 'borrowed from experience'. However, if Kant recognises such criteria, then the *a priori* status of the formula of humanity would not guarantee its application *a priori* to all human beings. As Charles Mills puts the point, if Kant's criteria for personhood are empirical, "what is a priori is that all rational beings are deserving of our respect; it is *not* a priori that all humans are rational beings."[13] If this is Kant's view, it would not be *a priori* that all humans are deserving of our (moral) respect; it might not even be true. I shall suggest that Kant has reasons in principle for rejecting empirical criteria for personhood or rationality. This is borne out in his brief

[13] Mills (2017: 101).

discussion of children's moral standing. Nevertheless, at several points he appears to take such criteria to be relevant to questions concerning moral status, the permissibility of certain kinds of treatment, and the extent of certain human beings' rights. Indeed, the exclusions briefly mentioned in the introductory section are all grounded in one way or another on apparently empirical judgements Kant makes about the rationality of different groups.

All philosophical accounts whose definitions of 'person' and 'human being' differ are likely to run into certain epistemic difficulties when determining which human beings are persons and whether any non-humans are persons. For instance, some accounts, such as that of Peter Singer, associate personhood with a cluster of empirical capacities – including intelligence, self-consciousness, and awareness of oneself as a temporal being – at least some of which arguably admit of degrees. For such accounts, there may be a difficulty in deciding on a threshold for personhood.[14] Kant, however, seems to face an especially acute epistemic problem that is peculiar to his theory and affects it in principle. This is because for Kant, our status as rational agents is intimately related to our possession of a free will. Indeed, the free will is identified with the capacity of a rational being to determine their own action: "A *will* is a kind of causality of living beings in so far as they are rational, and *freedom* would be that property of such a causality, as it can be efficient independently of alien causes *determining* it" (4:448). Kant's distinctive and radical form of compatibilism, however, turns on the idea that our freedom must be understood as *noumenal*. To reconcile his commitment to physical determinism with his commitment to (the possibility of) transcendental freedom, which

[14] There is disagreement in the literature regarding whether Kant is committed to such a "threshold concept" of personhood; see especially Allais (2016: 18). As will become clear shortly, the textual evidence is equivocal.

he takes to be a precondition of morality, Kant locates our freedom outside of the natural order.[15] He understands freedom as an 'Idea of reason', "a concept [...] which goes beyond the possibility of experience" (A320/B377). This means that our freedom can never be exhibited in our intuition: "Freedom [...] is a mere idea, the objective reality of which can in no way be presented in [...] any possible experience" (4:459; see also 5:69). Thus, the idea of a rational being is not itself an empirical one for Kant.

What, then, could the marks of rationality be? If freedom, and so practical reason, is never an object of experience, since our experience only contains spatiotemporal representations related to each other through necessary causal laws, it would appear that we can never directly perceive that other beings are rational in the relevant sense. There is something of an asymmetry here between the first and third person cases. In my own case, I must think of my actions as free, and Kant thinks this provides our concept of freedom with "objective reality at least in a practical respect" (5:44). His suggestion is that our very consciousness of the moral law and of the demands it places on us "provides a fact absolutely inexplicable from any data of the sensible world" (5:43) – what Kant calls the "sole fact of pure reason" (5:31), from which the reality of freedom can be deduced. I cannot evaluate this controversial claim here.[16] However, the relevant point for our concerns is that, as Ralph Walker has suggested, "the agent that must be free is (in the first place) *oneself*"; it is unclear how Kant's argument could establish that it is "possible for me to recognize *other* rational beings."[17]

[15] The relationship between freedom and the causality of nature is a central theme in most of the major texts of Kant's critical period, but see especially the 'Third Antinomy' (A444/B472-A452/B480; A532/B560-A558/B586) for his (negative) resolution of the conflict. For useful discussion, see Wood (1984); Allison (1990); Moore (2003a), 'Second Theme: Freedom'.
[16] See Allison (1990), Chapter 13, for a detailed exposition and qualified defence of Kant's deduction.
[17] Walker (2017: 217).

Perhaps in my own case I know – or at any rate, cannot help but think – that I *could* act morally. But as Walker points out, Kant is clear that we can never know with certainty that someone *did* act morally. As Kant emphatically puts it, "it is absolutely impossible by means of experience to make out with complete certainty a single case in which the maxim of an action that otherwise conforms with duty did rest solely on moral grounds" (4:407). The fact that an action conforms with duty in no way confirms that it was done *from* duty.[18] If we cannot know whether anyone else ever has acted morally, we might begin to wonder how we can know that anyone even has the capacity to do so. In Walker's words, "we could never 'know', in [this] limited sense, that someone was rational enough to be aware of the moral law and yet acted against it."[19] If no action counts as evidence that someone did act rationally (morally), what could count as evidence that someone *could* do so? Christine Korsgaard has suggested that for Kant, "human beings are distinguished from animals by the fact that practical reason rather than instinct is the determinant of our actions."[20] But if practical reason is never directly exhibited in experience, what grounds the claim that our actions (the actions of human beings generally) are *not* determined solely by instinct?

Of course, Kant does not endorse this sceptical line of thought. In most instances he assumes, naturally enough, that we can impute actions to other human beings – that we know that the moral law applies to them as it does to us, and that we are obliged to treat them as ends in themselves for that very reason. In response to the question he has posed for Kant, Walker's solution is to suggest that "we should treat as rational,

[18] This applies even in the first personal case, for Kant thinks that we are capable of great self-deception: our action could always have been the result of "a covert impulse of self-love under the mere pretense" of the idea of duty (4:407).
[19] Walker (2017: 218).
[20] Korsgaard (1996: 110-1).

and therefore as an end in itself, every being about whom we have *some* reason to think that it is rational."[21] This is because the law obliges us *never* to treat rational beings as mere means. The cost of failing to treat a rational being as rational is thus extremely high, and Walker seems to suggest that we should err accordingly on the side of caution. While he laments the fact that Kant did not state this point explicitly, he speculates that Kant may simply have thought it was obvious.[22] Some passages do support this speculation, if only indirectly. For instance, in a discussion of the rights of children to the care of their parents, Kant appears to suggest unequivocally that children are *persons*, and thus have moral standing:

> Children, as persons, have from procreation an original innate (not acquired) right to the care of their parents until they are able to look after themselves, and they have this right directly by law, that is, without any special act being required to establish this right. For the offspring is a *person*, and it is impossible to form a concept of the production of a being endowed with freedom through a physical operation. (6:280)

Children, then, are apparently 'endowed with freedom' from birth; as persons, they should thus be accorded certain rights.[23] This does not, however, settle the question of whether some human beings might nonetheless be 'beings without reason'. For Kant elsewhere seems rather more inclined to doubt the rationality of certain human beings, and thus their personhood in the fullest sense. Various remarks in Kant's moral and anthropological writings suggest that he does in fact countenance the idea that some human beings are not (fully?) rational. Consider the following distinction Kant makes

[21] Walker (2017: 218). Unlike Kant, Walker thinks that this point applies in the case of non-human animals (2017: 218-9).
[22] Walker (2017: 218).
[23] This curious passage seems to me to raise difficult questions regarding Kant's conception of freedom. If a child is endowed with freedom, why do we not treat them as fully culpable for their actions? But if the point is that they are not *yet* capable of acting freely, but will become so, how does this process take place? Are no 'physical operations', such as the various physical events that mark their transition to adulthood, involved in their becoming capable of acting freely? Unfortunately, however, I do not have the space to discuss these questions here.

between those with 'mental deficiencies' on the one hand, and those with certain forms of mental illness, namely 'mental derangements', on the other:

> The simpleton, the imprudent person, the stupid person, the coxcomb, the fool, and the buffoon differ from the mentally deranged not merely in degree but in the distinctive quality of their mental discord, and because of their ailments they do not belong in the madhouse; that is, a place where human beings, despite the maturity and strength of their age, must still, with regard to the smallest matters of life, be kept orderly through someone else's reason. (7:202)

The first group, those with mental deficiencies, are spared the madhouse. But the clear implication is that the particular 'quality' of mental derangement licenses us in keeping 'the mentally deranged' in 'a place where human beings must be kept orderly through someone else's reason'. While Kant does not go into further detail as to the character of this quality (or what about it justifies this treatment of human beings), the idea is presumably that their lack of reason justifies an imposition of reason onto their behaviour – they cannot keep *themselves* orderly through reason, so 'someone else' must ensure that they are kept in order. Whatever the precise justification is meant to be, it is evident here that reason is not something guaranteed to human beings *qua* human beings; if it were, no one would 'belong in a madhouse'. Furthermore, this passage suggests an interesting asymmetry: while Kant thinks that (moral) rationality cannot ever be inferred with certainty from one's behaviour (even if one acts in accordance with the moral law), he seems to think that a *lack* of reason can be manifest in one's behaviour in cases where that behaviour exhibits a particular kind of disorder.

In another passage, we are told that, owing to their 'immaturity' in civil matters, "women cannot personally defend their rights and pursue civil affairs for themselves, but only by means of a representative" (7:209). Kant understands immaturity to mean

the "(natural or legal) incapacity of an otherwise sound human being to use his[!] *own* understanding in civil affairs" (7:208). Women's immaturity is described by Kant as a 'legal immaturity,' rather than a natural one, which perhaps reveals an ambiguity in Kant's statements here: are they merely a description of the law as it stands, or are they supposed to carry normative weight? These are not, however, necessarily opposed. In Kant's account, they appear to come together. Just as parents are the "natural guardians" of their children, Kant thinks that a woman has a "natural curator" – namely her husband (7:209). The arrangements enshrined in the legal institutions of his time, Kant seems to think, are perfectly natural.

Kant occasionally made similar, if more extreme, comments with regard to race, writing for instance that "[Native] Americans and Blacks cannot govern themselves, so serve only for slaves" (15:878).[24] The reasoning here differs from the idea implicit in Kant's remarks on 'the mentally deranged'. There, Kant had suggested that a kind of paternalism justified the 'treatment' of those condemned to the madhouse – they cannot govern themselves, so someone else must bring order to their behaviour. Here, however, there is apparently not even the pretence of a paternalistic impulse. The alleged inability to govern oneself is made into a ground not only for someone else to govern 'on one's behalf', but for one's life to become completely subordinate to the ends of another; one now serves, can 'only' serve, that other as a slave.

In all three of these examples an idea of governance is operative, although the scope of that governance differs from case to case. The mentally deranged must be 'kept orderly'

[24] *Reflexion* 1520. This statement is an especially stark one, and comes from an unpublished note. As Allais points out, Kant does not actively endorse slavery in any of his published writings; see (2016: 5). Nonetheless, the fact that Kant could write a statement like this at all should presumably have some bearing on our interpretation of his published statements about racial differences and their significance. See Mills (2017: 97-8) for a sample of these.

'with regard to the smallest matter of life' – that is, their self-governance is replaced entirely and totally by the governance of another, by the administration of 'someone else's reason'. In a typical example of 'separate spheres' ideology, Kant allows women governance of the domestic sphere, at the cost of barring them from participating in their own governance at the political level. Finally, in the case of '[Native] Americans and Blacks', Kant infers from their supposed inability to govern themselves that they are natural slaves. Not only are they denied any sphere of self-governance, but their lives are made subservient to the ends of others. Such an inference surely evinces, among other things, an utter failure to regard and treat the members of these groups as ends in themselves.

As Charles Mills points out, it is hard to imagine "a more flagrant violation of the prohibition against using one's fellow-persons as mere means to an end" than the enslavement of a person.[25] Yet we find in Kant little in the way of explicit repudiation of slavery as a contemporary historical institution, and indeed some positive statements that certain groups of human beings are by nature fit to be slaves. This leads Mills to the exegetical claim that "when Kant urged on us the overwhelming importance of respecting persons, he was really talking (on this planet) about whites (more precisely, a subset of whites)."[26] In other words, Mills claims that Kant was operating with a view of personhood that (on his view) excluded a number of human beings from that category, on something like empirical grounds. Given that Kant appears to make distinctions of moral status based on determinations of race and gender, Mills suggests that "what is supposed to be the starkly polarized moral geography of his theory, with everything

[25] Mills (2017: 101).
[26] Mills (2017: 106).

being categorizable either as a *person*, with full moral status, or as a *non-person*, a *thing*, with zero moral status, would have to be redrawn to accommodate the fuzzier category of entities with some *intermediate* status" – hence, Mills' use of the term *Untermenschen* in his description of what he takes to be Kant's theoretical commitments.[27]

Now, the fact that the reasons Kant gives for excluding certain human beings from the scope of the formula of humanity are, broadly speaking, empirical generalisations – moreover, from our perspective at least, obviously false ones – might prompt the following response from the Kantian: "Why should it matter what *Kant* happened to believe about the rational capacities of particular groups of human beings? A more inclusive reconstruction of the Kantian view is a simple matter – after all, we now know that Blacks are perfectly capable of governing themselves, that women are perfectly able to participate in civil affairs without the need of a man to represent them, and so on. The empirical falsehoods Kant's own views were based on have been jettisoned, and the path is clear to a celebration of all human beings as ends in themselves."

While this would be a natural response, and one that seeks to preserve the appealing egalitarianism that is often associated with Kant's theory, I will argue that human beings with 'mental derangements' cannot easily be accommodated as members of the moral community on Kantian terms, even after we jettison Kant's other false empirical assumptions about who is to be counted as 'fully' rational.[28] More generally, the

[27] Mills (2017: 98).

[28] It is also worth noting that, as Mills points out, there are dangers involved in presenting reconstructions like the one sketched in the main text as straightforward exegesis of a philosopher's views. Indeed, Mills questions the extent to which we can really call an exegesis along those lines a 'Kantian' theory, if this covers up Kant's own commitments to racial hierarchies and contributes to "sanitiz[ing] and mystify[ing] the actual record of the past few hundred years by constructing the West as if white racial domination had not been central to

following worry will still remain: so long as moral standing is made contingent on the possession of a faculty that the Kantian thinks certain human beings may or may not possess, it will be difficult to guarantee the application of the formula of humanity to all human beings.

B. Ends in Themselves and the Objective Attitude: The Limitations of Kantian Perspectives on Mental Illness

The reason that this is important is that it is itself a moral question *which* creatures we regard as ends in themselves. This is not just a question of which beings receive a certain designation, but a matter of our entire attitude and practical orientation towards them. As we saw Walker point out, the cost of 'getting it wrong' in this domain is potentially very great. My contention in this section will be that Kant – and Kantian views more generally – draw too tight a connection between the notions of ends in themselves and of rational agency. Kant's moral theory leaves room for the idea that if some humans lacked the capacity to reason – if we could no longer locate their behaviour within the space of reasons – they could no longer be regarded as ends in themselves.

The notion of an end in itself is an enormously important one. It provides us with a vocabulary for articulating the wrong of certain forms of objectification, most significantly those forms that involve treating a living being as a tool or instrument. But there are more and less narrow ways of construing this notion. For Kant, only those creatures who are capable of freely determining their own ends have the status of ends in themselves. As we will see in Section II, such a conception makes it difficult to view

---

[its] history" (2017: 95). Nonetheless, Mills does not take these considerations "to rule out a revisionist anti-racist *appropriation* of [Kant's] thought" (2017: 230, fn. 46, emphasis added).

many non-human animals as ends in themselves. But it also presents difficulties in the human case. Even if we take Kant's sexism and racism to be merely incidental to his moral theory, and hence easily overcome by its modern descendants, an analogue of Kant's treatment of 'mental derangement' arguably persists. 'Extreme' forms of mental illness still appear to be taken by Kantian philosophers to constitute a kind of limit to the attribution of rationality; 'deranged' human beings are located outside the space of reasons, all but licensing a denial of their status as ends in themselves and an exclusion of them from the Kantian 'kingdom of ends'. By attending to these cases and their 'treatment' by Kantian authors, I will argue that this suspension of moral status is at best premature, and at worst itself actively objectifying. The upshot of this is that if we are to acknowledge the moral claims made on us by all human beings, we should not accept the Kantian claim that rationality is what accords us dignity and moral status.

To provide the necessary background to this claim, I shall begin by examining some expositions of a distinction, originally drawn by Peter Strawson, that is closely related to Kant's distinction between persons and things. Unlike the latter, however, Strawson's distinction is not between two kinds of entity, but rather between two kinds of attitude that we take up towards entities. These are the 'participant attitude', "the attitude (or range of attitudes) of involvement or participation in a human relationship", and the "objective attitude (or range of attitudes)", which Strawson explains as follows:

> To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided, though *this* gerundive is not peculiar to cases of objectivity of attitude.[29]

[29] Strawson (2008: 9).

For simplicity's sake, I shall follow Strawson in talking about *the* objective attitude in the singular, though as the passage just cited makes clear, Strawson intends us to understand the term in a capacious way, with several distinguishable strands. Strawson's description makes clear that while the objective attitude is a way of *seeing* a human being *as* such-and-such, this way of seeing has immediate ramifications regarding how we are inclined to behave towards that human being. Indeed, perhaps one reason 'objective' is an apt term here is that we are inclined to treat the human being to whom we adopt the objective attitude as in certain respects object-like: we will not treat them as a subject to be reasoned with, but as some*thing* to be dealt with or manipulated in a variety of ways.[30] Another possible resonance of 'objective' is its connotation of a measure of detachedness. Strawson suggests that the attitude arises primarily where we are inclined to suspend our ordinary reactive attitudes (such as anger, resentment, gratitude etc.) to the actions – or apparent actions – of a particular human being. It arises when we no longer feel as if these responses are appropriate, usually because for some reason or other we do not deem the human being in question to be responsible for what they have done.[31] It is worth noting that these strands come apart: even if I suspend attitudes of blame, praise etc. and do not hold someone as responsible for their actions, this does not *yet* mean that I will be willing to treat them as merely 'to be managed', 'to be taken precautionary account of', and so on.

Rae Langton has characterised the participant attitude – the ordinary interpersonal standpoint from which Strawson distinguishes the objective attitude – as an "interactive" standpoint: "When you hold someone responsible, you are prepared to

[30] Strawson (2008: 10).
[31] See e.g. Strawson (2008: 20-1), where Strawson glosses "a responsible being" as "an object of non-objective attitudes".

work with them, view them as someone who has goals of their own that you might come to share, or as someone who might come to share your goals. You are prepared to *do something with them*, in a sense very different from the sense in which you might do something with a tool."[32] She illustrates this point with the example of baking a cake with a friend. In such an activity, one *does something with* the friend (namely the shared activity of baking). At the same time, but in another sense, one *does something with* the ingredients, the wooden spoon, the mixing bowl, and so on – the 'tools' and 'things' one manipulates in order to produce the finished cake. The difference is that the friend is genuinely sharing in one's activity – in a way that the tools cannot – because she "can choose ends of her own, and can choose to make them coincide with mine."[33] In this way, a distinctively Kantian understanding of the participant attitude emerges. Langton herself relates Strawson's discussion explicitly to Kant's, drawing out the Kantian background that was only implicit in Strawson's original article. For both Strawson and Langton, 'involvement and participation in a human relationship' is in large part a matter of human beings, who each stand in a free relation to their own ends and recognise each other as such, choosing to interact with each other. Thus, Langton relates her understanding of the participant attitude directly to Kant's distinctions between person and thing, dignity and price: "The standpoint we take towards human beings is interactive, and it is different from the standpoint we take with things. Kant thinks this is because human beings have an intrinsic worth that has its basis in our capacity for rational choice. Human beings are ends in themselves, who have a dignity, and not a price."[34]

[32] Langton (2009a: 204).
[33] Langton (2009a: 204).
[34] Langton (2009a: 204).

There is certainly something true and important about this conception of the participant attitude. Treating someone else as an end often means recognising their capacity to form their own ends – what Kant would call their autonomy, their practical rationality. Moreover, as Langton has argued compellingly, *denying* someone's autonomy – whether it be through non-attribution or violation of their autonomy – is a way of failing to treat them as an end.[35] It is a form of objectification. It is thus unsurprising that Langton at one point equates the interactive standpoint with "the moral standpoint."[36] Nevertheless, there are reasons for questioning whether this Kantian conception of what the participant attitude involves adequately covers the whole range of meaningful, non-objectifying relationships that human beings may be involved in – with each other, and indeed with non-human animals.

Consider first a variation on Langton's baking example: suppose a parent is baking with their young child. Perhaps the child can perform a certain number of the tasks involved in baking, but not others: they can stir the mixture, or hold the sieve for the flour, but cannot yet measure out the ingredients. The child might understand some simple instructions, but not others; the parent might make a game of the activity to help them learn. The child might express delight, or again boredom. Now, in which of Langton's two senses is the parent 'doing something with' the child? Obviously not in the sense that they are 'doing something with' the ingredients and utensils. The child is participating in the shared activity of baking, albeit in a more limited way than an adult would be. The appropriate attitude to take towards them is a special mode of or variation on the participant attitude. However, it is not obvious that Langton's gloss on

[35] See especially Langton (2009b).
[36] Langton (2009a: 204).

what that attitude amounts to adequately captures the texture of this specific human relationship. To a great extent, the child may not (yet) be able to choose or reflect on their ends, and to make them coincide with those of the parent.

It is significant that this does seem to be a question of extent or degree, and Strawson's talk of "the progressive emergence of the child as a responsible being" seems apt.[37] Nonetheless, I think something goes phenomenologically awry when Strawson suggests that this is also an emergence of the child "as an object of non-objective attitudes", and when he suggests that "parents and others concerned with the care and upbringing of young children cannot have to their charges either kind of attitude in a pure or unqualified form."[38] The attitude of a parent to their young child is of course different in many respects from the typical participant attitude that one adult might hold towards another, but those differences need not make the former attitude any more 'objective' than the latter. As some of Strawson's own remarks suggest, the parent sees the child not so much as 'to be managed or handled or cured or trained', as 'to be *cared for*' or 'to be *nurtured*'.

Perhaps a still more difficult case to locate is that of non-human animals, especially companion animals of various kinds. Some elements of Strawson's characterisation of 'the objective attitude' might seem more apt here than they did in the case of the child: one *trains* a dog or a horse, perhaps *handles* them, and if they become dangerous to human beings or other animals, one 'takes precautionary account' of them. But consider a variation on the question I asked earlier: in what sense is someone 'doing something with' the animal when they ride a horse, or are aided by an assistance dog?

[37] Strawson (2008: 20-1).
[38] Strawson (2008: 21, 20).

In both cases, it would seem to be a gross mischaracterisation of the relationship between human and animal to view the animal as akin to the cooking utensils and cake ingredients – as a 'mere means' to an end, as a tool or implement that is not itself participating in the activity and possessed of a perspective on it.[39] (It is not irrelevant here that in both of my examples the animals are doing most of the heavy lifting, so to speak.) But again, it does not seem quite right to say that these animals are able to choose their ends, and choose to align them with ours – at least not if 'choose' means something like 'rationally reflect on'. Horses and dogs are not autonomous in the special Kantian sense. But for all that, I am tempted to say that they are free in their own way.[40] Furthermore, it seems clear that in the case of non-human animals we can also distinguish between more and less interactive, more and less objective attitudes. Animals are also often wronged by forms of objectification, most obviously when they are slaughtered, packaged up, commodified, and consumed.

The upshot of all of this is that the distinction between the 'participant' and 'objective' attitudes is rather more complicated than it might at first appear. In particular, there is perhaps a danger of building a theory-laden conception of what a properly human relationship looks like – or should look like – into the characterisation of the participant attitude. To return to my main topic, I want to draw attention now to a way in which such a characterisation might seem to license a pernicious exclusion of some human beings from the sphere of moral consideration – in other words, from the sphere of beings to whom it would be appropriate (or indeed obligatory) to take up a non-objective attitude. For there are clear affinities between Strawson's description of

[39] As we shall see in the next section, the principal failing of Kant's own treatment of animals – perhaps in multiple senses of 'treatment' – was to view them precisely in these kinds of terms.
[40] Cf. the closing paragraph of Srinivasan (2021).

the objective attitude and what Kant had to say about 'mental derangement'. Seeing someone as 'to be managed or handled or cured or trained' is not altogether different from seeing them as to 'be kept orderly through someone else's reason'. Further, it is clear from Strawson's discussion that he takes the "mentally deranged" human being to be someone toward whom we paradigmatically take up the objective attitude.[41] The same association between the objective attitude and mental illness is suggested by an illustration Langton provides of the former:

> Suppose my neighbour forms a habit of vandalizing my car when it is parked outside his house. He emerges stealthily, at dead of night, and gently twists the wipers into intricate and elegant knots. The next day I stride to his door and knock, brimming with indignation, planning to ask him to be reasonable, hoping to reach mutual understanding. But if he responds with bulging eyes and a torrent of incoherent invective, and I see that he is a badly shell shocked war veteran, indignation instantly vanishes, to be replaced by pure alarm. I stop thinking of him as an agent, whose reasons, mysterious as they might be, I can in principle come to understand. My neighbour becomes a problem to be managed, an obstacle to be avoided, not a person to be argued with. He becomes just one more of the hazards of Elwood, along with the threat of the flooding canal. I have switched from the participant standpoint to what Strawson calls the *objective*. This is the attitude we have to things, items in the natural order, whose behaviour is explicable under causal laws, and manipulated if you know enough about them.[42]

Langton's narrative very clearly echoes the terms used in Strawson's original characterisation of the objective attitude. After the shift in standpoint, the neighbour is no longer seen as a person with whom one might reason or 'argue', and becomes a problem 'to be managed', 'to be avoided'; the understandable interpersonal response to their initial behaviour ('indignation') is suspended, 'vanishes', to be replaced by a less personal affective state ('pure alarm'), one that might equally appropriately be evoked

[41] See e.g. Strawson (2008: 9, 10, 13).
[42] Langton (2009a: 204-5). I am grateful to Sophie Russell for bringing this passage to my attention and discussing it with me.

by an impersonal danger (such as the 'threat of the flooding canal' with which the neighbour is compared and rhetorically equated).

Langton's discussion might seem to imply that the objective attitude is only objectifying when it involves a genuine denial of autonomy – when the subject of the attitude could be understood as autonomous, but is treated as if they were not. If a being is not autonomous (in the Kantian sense) to begin with, if they are "deranged or compulsive", then taking up the objective attitude towards them is not only appropriate but "seems to be forced upon us".[43] They are simply another part – another potentially threatening part – of the natural world, to be manipulated accordingly. But surely Langton's narrator takes up too extreme an objective attitude in making this final judgement, in seeing the neighbour as really on a par with the causally explicable natural order. (Incidentally, faced with this appeal to the 'natural order', one might be tempted once more to raise the uneasy question of situating non-human animals in relation to Strawson's distinction between standpoints.)

Presumably, the scenario is intended to be one where the reader is likely to judge the objective attitude to be appropriate. But it is not entirely clear whether either Strawson's or Langton's discussions are supposed to be simply descriptive of the attitudes we do in fact (sometimes) take up, or normative, providing something like a justification for our adoption of the objective attitude towards certain human beings, or an account of when that attitude is appropriate. However, Strawson does suggest that the objective attitude is the "civilized" attitude to take up towards those who are

[43] Langton (2009a: 205); see also Strawson (2008: 10).

"warped or deranged or compulsive in behaviour".[44] Presumably a 'civilized' attitude is meant to contrast here with an 'uncivilized' one, though Strawson does not elaborate on what such an attitude towards the 'deranged' person might amount to. Given the context, his point may be just that it would be inappropriate to view them as responsible for their behaviour, to treat them as morally blameworthy for the effects of their illness. This much seems plausible. But more seems to be involved in the objective attitude (as Strawson describes it) than just the suspension of attitudes of blame, praise, etc. This affective strand of the attitude comes apart from the question of our positive treatment of a human being whom we do not regard as fully responsible for their actions. It also comes apart from the question of what kind of viewpoint they might have on themselves and on us.

This brings out a further aspect of the objective attitude. Langton's discussion seems to portray the mad neighbour as someone who has no viewpoint at all – or at best something like an animal viewpoint. If we do not understand the case in this way, then the equation between the mad neighbour and a mere 'item in the natural order' becomes not just hyperbolic, but objectifying – a form of dehumanisation. It is difficult to imagine how such an extreme form of the objective attitude could be sustainable on reflection. Nonetheless, Strawson also appears to view this attitude as an appropriate one, and repeats his claim that it is the only 'civilized attitude' available in such cases:

> In the extreme case of the mentally deranged, it is easy to see the connection between the possibility of a wholly objective attitude and the impossibility of what we understand by ordinary inter-personal relationships. Given this latter impossibility, no other civilized attitude is available than that of viewing the

[44] Strawson (2008: 10; see also 13).

> deranged person simply as something to be understood and controlled in the most desirable fashion.[45]

Presumably the sense of 'understood' mentioned in the final sentence of this passage is supposed to refer to a purely objective, causal understanding comparable to what Langton refers to – a form of understanding that we would usually take to be more appropriate to the 'objects of nature'. It is not an inter-personal or involved understanding – the kind of understanding where we try imaginatively or empathetically to 'see where someone is coming from', to understand their standpoint or perspective on things. But even in the 'extreme case of the mentally deranged', there is surely still a perspective there that one has an obligation to try to understand. It is hard to imagine behaviour that would really 'force' us to view another human being as radically unintelligible in principle, so as to make such an effort otiose. Even in Langton's example, there remains a certain dissonance between the purposiveness – even the deliberateness – suggested by the adjectives and adverbs in her initial description of the neighbour's behaviour ('stealthily', 'gently', 'intricate', 'elegant') and the impression of mindlessness made on the narrator by his 'torrent of incoherent invective', his 'bulging eyes', and whatever it is about his appearance that allows her to infer that he is a 'badly shell shocked war veteran'. This is, of course, a fictional scene, and an intentionally surreal one. But what would it be like to try, imaginatively, to reconcile its elements? How would the world appear to someone of whom this description were true? What trauma might they have endured to bring them to this depersonalized – or all too easily depersonalizable – state? Could they be consoled or comforted – if they need consolation or comfort? How should we 'treat' them – in a 'wide range of sense'?

[45] Strawson (2008: 13).

I do not intend these to be merely rhetorical questions. They are both real and difficult. It is highly doubtful that they could be settled from the philosopher's armchair, rather than in actual, painful, and fallible human interactions. Whether I am right about this or not, Strawson's verdict on the final question seems to me much too casually delivered. It seems to close off the vast and ambiguous space between 'ordinary inter-personal relationships' and 'the wholly objective' (here, surely, object*ifying*) attitude. The word 'ordinary' appears to play a role in Strawson's description that is at once tacitly normative and highly questionable. Consider, for instance, the relationship between the psychiatrist and their patient. It is surely not an 'ordinary' relationship, but for all that it is – or ought to be – a relationship between human beings. Furthermore, where a patient is judged to be unintelligible, their supposed unintelligibility is not necessarily the result of their behaviour alone: as R. D. Laing puts it, it is also "created by the person diagnosing" them.[46] In such situations, the decision to treat another human being as unintelligible may effectively confirm and reinforce itself. Laing continues: "The worst violence of all is the reciprocal denial of reciprocity [...] both by the patient who refuses to communicate, and by the psychiatrist, who double-stamps this *refusal* as *inability*."[47] What measure of agency they have, which they exercise in refusing to communicate, is effectively concealed by the diagnosis that they are capable neither of controlling themselves nor of making themselves understood.

Such cases suggest that we do not have a *clear* idea of when it is really appropriate to adopt an unqualifiedly objective attitude, if it ever is. They also highlight the risk of pretending to ourselves that such an attitude is forced on us, in cases where we may

[46] Laing (2015: 18).
[47] Laing (2015: 20).

actually be adopting it "as a refuge [...] from the strains of involvement."[48] In such cases, a failure of imagination will also be a moral failure.

The upshot of all this is that the prospects for reconstructing Kant's formula of humanity in a way that protects the rights of all human beings look fairly dim, if we maintain the close ties between rationality and moral worth that Kant himself stressed. One response would be to suggest that the problem is not with the idea of rationality as such. Rather, the problem is that the conception of rationality endorsed by Kant (and, following him, Strawson and Langton) is too narrow. If we expanded our conception of rationality – perhaps to include the various forms of agency exhibited by children and the patients discussed by Laing – we could maintain a connection between rationality and the moral standing of human beings. I am not ruling out such a project here. However, I take it that such a project would involve a decisive break with Kant's conception of moral standing. For this was tied up not just with rationality, but specifically with our capacity for morality.

As we shall see, a similar problem arises for Christine Korsgaard's attempts to give a Kantian explanation for the moral standing of non-human animals. Before examining Korsgaard's arguments, I will briefly discuss Kant's own remarks about other animals, and the severe limitations of his account.

II. Animals as Things? Anthropocentric Strands in Kant's Humanism

A. Kant's Own View: Indirect Duties and Animals as Mere Means

At the beginning of this chapter, I briefly summarised Kant's account of our duties to non-human animals – or, better, our duties regarding the treatment of non-human

[48] Strawson (2008: 10).

animals. For as I suggested there, on Kant's view we do not have, strictly speaking, any duties *to* non-human animals. Instead, our treatment of them, where it is impermissible, is so because of duties we have to ourselves, to mankind in general. The quotation from the *Lectures on Ethics* is by no means an aberration in Kant's ethical writings. Kant offers a similar explanation of our duties "with regard to" but not *to* animals in the *Metaphysics of Morals* (6:443). Beyond these two brief discussions, Kant rarely mentions our duties (with regard) to animals. Still, his estimation of the negligible moral standing of animals is made abundantly clear in the scattered references to animate nature throughout his writings.

A recurring theme in these discussions is Kant's view that, unlike people, non-human animals can permissibly be treated as mere means, subject to the restriction – to be discussed presently – that they ought not to be treated cruelly (see e.g. 6:443, 8:114). In Section I, I discussed Kant's stark distinction between 'person' and 'thing'. It is clear that for Kant, animals fall on the latter side of this divide. In fact, this is obvious from the very way in which he draws the distinction: "beings the existence of which rests not on our will but on nature, *if they are beings without reason*, still have only a relative worth, as means, and are therefore called things" (4:429, emphasis added). The morally relevant distinction – what makes the difference between a person and a mere thing – is the distinction between beings with reason and beings without it. Kant does not deny that animals have sensations, asserting in a footnote that "nonrational creatures feel only sensible impulses", in contrast to rational creatures who can properly speaking "[take] an interest in something" (4:460). But this difference between nonrational creatures and other, non-animate 'things' is, morally speaking, immaterial (or *almost*

immaterial) on Kant's account. Kant's view, then, is not just anthropocentric, but profoundly so.

Christine Korsgaard has recently questioned the coherence of Kant's own account of animals' moral status. To understand the force of her criticism, consider a passage from the *Metaphysics of Morals* where Kant elaborates somewhat on his indirect duties view, and on the content of our duties with regard to animals:

> With regard to the animate but nonrational part of creation, violent and cruel treatment of animals is far more intimately opposed to a human being's duty to himself, and he has a duty to refrain from this; for it dulls his shared feeling of their suffering and so weakens and gradually uproots a natural predisposition that is very serviceable to morality in one's relations with other people. The human being is authorized to kill animals quickly (without pain) and to put them to work that does not strain them beyond their capacities (such work as he himself must submit to). But agonizing physical experiments for the sake of mere speculation, when the end could also be achieved without these, are to be abhorred. – Even gratitude for the long service of an old horse or dog (just as if they were members of the household) belongs *indirectly* to a human being's duty *with regard to* these animals; considered as a *direct* duty, however, it is always only a duty of the human being *to* himself. (6:443)

According to Korsgaard, Kant's view encourages an incoherent attitude towards animals. They are clearly different from *mere* things, since kindness and gratitude are not appropriate or even comprehensible attitudes to take toward the latter. The indirect duty view is therefore somewhat unstable, and tends towards the recognition that we really must owe something *to* the animals.[49] Furthermore, the very fact that cruelty to animals is something bad *for them* suggests that it make sense to think of them as having a final good, and thus a form of moral status.

So far, these are somewhat minimal points, compatible with a range of views on how much we owe to animals. But they nonetheless expose the extremity of Kant's own

[49] See Korsgaard (2018), Chapter 6, especially 6.3.

view. For even those who deny that our obligations to animals are very strenuous tend to think that they are obligations *to* those animals. Moreover, Korsgaard also supplies a more ambitious positive argument for regarding and treating animals as ends in themselves, which I shall now consider in detail.

B. A Formula of Animality? Korsgaard on Animals as Ends in Themselves

Although she sharply criticises Kant's own views regarding our obligations to other animals, Korsgaard nonetheless thinks that Kant provides us with resources that can make sense of those obligations. It is relevant for my purposes here that it is important to Korsgaard that her account be a distinctively Kantian one – that is, one that coheres with her other broadly Kantian commitments (viz. constructivism and the crucial notion of ends in themselves) and that proceeds from recognisably Kantian premises. Indeed, Korsgaard's argument is to my knowledge the only sustained attempt to counter Kant's own anthropocentrism from within his system.[50] Its success would vindicate the possibility that Kantian humanism could avoid anthropocentrism. Its failure would not, of course, show that such a view could not be formulated and defended. But it might suggest that the prospects for such a view look dim. It might also help us understand why a theory that starts from a broadly Kantian outlook is unlikely to be conducive to the project of making sense of our obligations to animals. With these possibilities in mind, I turn now to discussion of Korsgaard's argument.

Korsgaard's strategy is to argue that we are obliged to treat not just ourselves, but also non-human animals, as ends in themselves and therefore as having moral standing and

[50] With that said, other views that take non-human animals to have inviolable rights could also be described as 'Kantian' in a looser sense. See for instance the discussion of 'Kantian' and 'utilitarian' approaches to animals' rights in Donaldson and Kymlicka (2011: 20; see also 27).

as worthy of a certain kind of moral respect. By her own admission, the argument takes some "complicated twists and turns" (no doubt, as she herself suggests, an occupational hazard for scholars of Kant), but she also offers "a way to make [the argument] simply."[51] I shall begin by quoting the 'simple' version of the argument in full, and then proceed to discuss its most contentious premises in the light of Korgaard's earlier, more complicated exposition. The argument is as follows:

> [1] As rational beings, we need to justify our actions, to think there are reasons for them. [2] That requires us to suppose that some ends are worth pursuing, are absolutely good. [3] Without metaphysical insight into a realm of intrinsic values, all we have to go on is that some things are certainly good-for or bad-for us. [4] That then is the starting point from which we build up our system of values – we take those things to be good or bad absolutely – and in doing that we are taking ourselves to be ends in ourselves. [5] But we are not the only beings for whom things can be good or bad; the other animals are no different from us in that respect. [6] So we are committed to regarding all animals as ends in themselves.[52]

Premise [1] is straightforward enough. On Korsgaard's Kantian picture, to be a rational being *just is* to move within the space of reasons – this is what makes the difference between rational action and instinctual action.[53] We are able to step back from our actions, and interrogate our reasons for them. Premise [5] is also unproblematic, given a broadly naturalistic reading of the notion of something being 'good-for' or 'bad-for' some creature.[54] However, in premises [2]-[4] we run into difficulties. For Korsgaard's argument that animals must be treated as ends in themselves relies on the idea that we

---

[51] Korsgaard (2018: 145).
[52] Korsgaard (2018: 145, bracketed numbers added).
[53] See Korsgaard (2018: 38-44). Korsgaard also makes a further distinction between intelligent and rational action, suggesting that some non-human animals act intelligently but (as far as we know) only human beings are capable of acting rationally. This further distinction need not concern us here, since Korsgaard's argument is supposed to establish that *all* animals – all creatures with a final good – are ends in themselves, whether they act merely instinctually or intelligently as well.
[54] Indeed, given how Korsgaard defines 'creature', it is practically true by definition; see (2018: 16-7).

must treat that which is 'good-for' or 'bad-for' some creature (whether a human being or a non-human animal) as good or bad absolutely. It is not at all clear that such an idea would be acceptable to Kant. Before returning to the question of the status of non-human animals, I shall try to clarify this point by unpacking the notions of 'absolute good' at play in Kant and Korsgaard's work.

Readers familiar with Kant's *Groundwork* might balk at Korsgaard's use of the phrase 'absolutely good' here. For Kant famously claims there that the *only* thing that is good absolutely is the good will. In the first line of the *Groundwork* proper, he writes that "it is impossible to think of anything at all in the world, or indeed even beyond it, that could be considered good without limitation except a **good will**" (4:393), and the ensuing discussion makes it clear that this notion of 'good without limitation' coincides with Kant's conception of what is 'absolutely good'. The two phrases are near synonyms. In discussing the question of the possible role of empirical additions to the principle of morality, and anticipating the distinction between price and dignity that he makes later in the *Groundwork*, Kant writes that "the proper worth of an absolutely good will – a worth raised above all price – consists just in the principle of action being free from all influences of contingent grounds, which only experience can furnish" (4:426). As we saw earlier, in that later discussion Kant proceeds to equate the notion of an end in itself with that which "has not merely a relative worth, that is, a price, but an inner worth, that is, *dignity*" (4:434). In Kant's view, then, it appears that the dignity of ends in themselves is essentially related to their freedom from the influences of contingent grounds, and that this in turn is intimately connected with the possibility that their will can be absolutely good, or good without limitation.

If Korsgaard is using the term 'absolutely good' in Kant's sense, then, we can draw out two *prima facie* difficulties for this part of her argument. The first is the idea that 'absolutely good' is an acceptable gloss on 'worth pursuing'. If only one thing in the world is good without limitation, absolutely good, then presumably the majority of ends that we pursue are merely relatively or conditionally good. But they may be, for all that, worth pursuing – worth pursuing, that is, given some condition. This brings us to the second difficulty, which is with premises [3] and [4]. It is not clear how, on Kant's view, we can take what is 'good-for' and 'bad-for' us as the basis of determining what is good or bad absolutely. For one thing, the concepts of what is good-for or bad-for us are furnished empirically. Our own happiness or well-being is an empirically given end, even if it is one that we can assume each creature possesses. It is never an absolute good in Kant's sense.

Therefore, Korsgaard must mean something different by 'absolute good' than what Kant meant by it. To understand this part of the argument, let us turn then to her explanation of this term. One of her starting points is the idea that importance and goodness are "tethered notions", which means "that nothing can be important without being important-to someone, and nothing can be good without being good-for someone."[55] A natural worry about such a conception would be that it relativizes the idea of goodness – one might wonder what standpoint there is left from which we can say that something is good absolutely, or good *simpliciter*. Korsgaard's reply to this worry explains her conception of what it is for something to be absolutely good: "When we say something is good absolutely, what we mean is not that it has a free-floating goodness, but that it is good-for everyone for whom things can be good, in the final

[55] Korsgaard (2018: 135); see also Chapter 2.

sense of good, or good from everyone's point of view." She then suggests that we can put the same point another way: that which is good absolutely "can be included in a universally shared or common good, one that we can all pursue together." One benefit of this second formulation, according to Korsgaard, is that "it will enable us to include things that are good for someone, so long as they are not bad for anyone, among the absolute goods." Summing up this discussion, Korsgaard writes: "to say that something is good absolutely is just to say that it is good-for us all – that is, good (or at least not bad) from every point of view, part of a universally shared or common good."[56]

In certain respects, Korsgaard's conception of absolute goodness resembles Kant's own. In particular, the idea of something that is 'good from everyone's point of view' arguably coincides with Kant's idea of something that is good without limitation: that which is only good from a particular perspective is good in a limited or conditional way, but that which is good from any possible perspective has no such limitation. Kant's argument for the formula of humanity, at least on Korsgaard's reconstruction, relies on precisely this idea: to value anything at all, I must value my own ability to value, my capacity to set ends – in other words, I must value my rational nature. But there is nothing special about *me* in particular – rather, it is rational nature as such that I must value. So, I must treat rational nature 'wherever I find it' as an end in itself. Anyone who values anything at all is therefore committed to valuing rational nature.[57]

Suppose that argument goes through. It would still be far from clear that anyone capable of valuing things at all must value anything that is good-for someone, as long as it is not bad for anyone. In other words, the second formulation that Korsgaard offers

[56] Korsgaard (2018: 136-7).
[57] See Korsgaard (1996: 123).

is not obviously equivalent to the first. For it is simply not true that everything that *can* be included in a universally shared or common good that we *can* all pursue *must* be included in that universally shared good. Something can harmonise with the common good without everyone regarding it as good – without it being good from everyone's point of view. I might value long-distance running, and adopt as my end the resolution to train six times a week. But I do not regard long-distance running as worthwhile from everyone's point of view – I regard it as worthwhile from *my* point of view. Furthermore, your respect for my choice need not – I am tempted even to say that it *must* not – depend on your also regarding my end as worthwhile. What grounds your respect for my choice is the fact that it was *my choice*, rather than anything about the content of that choice.

This is connected with Korsgaard's limitation on the ends that we can consider absolutely good: '…so long as they are not bad for anyone.' (The very fact that there can be such a limitation further demonstrates the distance between Kant's conception of 'absolute good' and Korsgaard's.) If we adopt Korsgaard's terminology, leaving aside for a moment the ends that are bad-for someone, among the ends that are left it is unclear what distinction can be made between that which is absolutely good (good from everyone's point of view) and that which is relatively good (good from someone's particular point of view). This seems to me to be a bizarre result. Respecting someone as an end in itself means respecting their ends even when you do not think that those ends are worthwhile, so long as those ends are not bad-for anyone – that is, so long as they do not contravene the moral law. The ends themselves are not absolutely good; it is the capacity to choose those ends that is absolutely good and that must be respected, whatever one's point of view.

This comes out in a particularly striking way in Korsgaard's discussion of our natural ends. Korsgaard suggests that we value ourselves not only as rational beings, but also merely as animal beings. She claims that this fact "becomes especially clear when we reflect on the fact that many of the things that we take to be good-for us are not good for us in our capacity as autonomous rational beings. Food, sex, comfort, freedom from pain and fear, are all things that are good for us insofar as we are animals."[58] This is no doubt true. But are any of these goods absolute in anything like Kant's sense? It is not clear that these are acceptable as Kantian premises. The inclusion of 'sex' under the list of the things that are good for us would no doubt have alarmed Kant himself, but such a consideration is obviously not decisive; Kantian ethics is more concerned with what Kant should have said about such-and-such, given his deeper theoretical commitments, than with what he actually said. In any case, there *is* a deeper theoretical commitment here, which is the question of what role happiness or well-being should play in moral thought. Kant's view is clear: happiness is a good worth pursuing, but it is a conditional good. It is absolutely subordinate in value to the good will, motivated by the moral law:

> This [good] will need not [...] be the sole and complete good, but it must still be the highest good and the condition of every other, even of all demands for happiness. [...] It is entirely consistent with the wisdom of nature if we perceive that the cultivation of reason, which is requisite to the first and unconditional purpose [viz. to produce a good will], limits in many ways – at least in this life – the attainment of the second, namely happiness, which is *always* conditional. (4:396, emphasis added)

It is surely an overstatement to say, with Korsgaard, that what is good-for and bad-for us is *all* we have to go on when we set ends for ourselves. We also have access to the pure moral law, which is given in abstraction from all empirical conditions. It is true that we value ourselves as animals, that we act on the basis of our animal needs. But that which

[58] Korsgaard (2018: 144).

is good-for us as animals cannot be absolutely good. It is only good insofar as it is permitted by the moral law; even our duty of beneficence, of increasing the happiness of others, is so limited. Korsgaard is surely wrong to claim that on Kant's view, "beneficence requires respect for someone's animal nature, not merely for his rational nature."[59]

So far, I have proceeded by criticising Korsgaard's argument on what I take to be broadly Kantian grounds. However, my criticisms have not yet addressed a further crucial part of the argument, which is the step from valuing my own ends absolutely to valuing the ends of all creatures absolutely. Korsgaard writes, "I am deciding to treat my ends as good absolutely, simply because I am a creature with a final good. From there all we have to do is generalize: that principle requires that we should take the ends of beings who have a final good to be absolutely valuable."[60] I now want to make a final point regarding this apparently simple generalization. Kant believed that we could hope for – if not, in this life, achieve – a state of affairs in which all ends are harmonized with each other, where we can all pursue our ends without infringing on the ends of others. This is his idea of the Kingdom of Ends. But the difficulty, as Korsgaard herself puts it, is that "once we invite the animals in to the Kingdom of Ends, that hope of making the world good for everyone is gone."[61] What is good for the lion is not good for the antelope, and vice versa.

Korsgaard suggests that while the impossibility of complete harmony among the ends of animals presents us with many difficult questions, it does not undermine the idea that we should treat animals, when we encounter them, with the respect due to ends in

[59] Korsgaard (2018: 145).
[60] Korsgaard (2018: 144).
[61] Korsgaard (2018: 154).

themselves: "The ineradicable conflicts in nature are *in general* no reason for human beings not to treat animals as ends in themselves."[62] For instance, the issue of whether we can advance the ends of the lion and the antelope at once has little bearing on the question of whether we have a duty not to raise animals in factory farms for human consumption. But granting that point, it seems to me that this difficulty does raise a devastating problem for Korsgaard's procedure of generalization. To see why, recall the restriction on including things that are good-for someone among the absolute goods: private goods can be included "so long as they are not bad for anyone." But this is not and perhaps cannot be true of the natural ends of animals. The lion's good is bad for the antelope (the antelope will be consumed); the antelope's good is bad for the lion (the lion will go hungry).

If we accept the restriction, it is not clear how we can accept the generalisation. But if we reject the restriction, we will be committed to thinking, absurdly, that private goods that are *bad* for some creatures can be included among the 'absolute goods'. Either way, it is not clear how Korsgaard's argument can succeed. I want to emphasise, however, that I am not advancing this criticism to cast doubt on the idea that we have obligations to non-human animals. On the contrary, the point is only that it is not at all clear that the Kantian idea of the 'Kingdom of Ends' can adequately ground and make sense of those obligations. A moral discourse centred on rational agency, and its corollary ideal of a moral community of self-legislating agents pursuing their ends in harmony, seem to me singularly ill-fitting as a basis for moral thought about the relations between human and non-human animals.

[62] Korsgaard (2018: 155).

Conclusion: Beyond Kantian Humanism?

I hope to have shown that it is far from clear whether we can reach a non-anthropocentric humanism from Kantian premises. Indeed, it is unclear whether we can make sense of our obligations to either non-human animals or to those human beings who, for whatever reason, lack the capacity for rationality, if we remain within the Kantian framework. Nonetheless, we should not necessarily jettison the notion of an end in itself, even if we reject the grounding Kant gave to it. Even if Korsgaard's Kantian argument for treating non-human animals as ends in themselves fails, it by no means follows that we are thereby warranted in treating them as mere means. The critical question is how we should make sense of the idea that there is something wrong with treating creatures – human and non-human – as mere means, even in cases where we have grounds for viewing those creatures as 'beings without reason'.

Chapter Two
Thinking Against Humanism?
Heidegger on the Human Essence, the 'Inhuman', and Evil

In one way or another, the question of the human being, of its 'existence' and 'essence' (though not always in the familiar metaphysical senses of these terms), is of central concern to Heidegger in both his early and later works. These questions are given sustained treatment in the *Letter on 'Humanism'*, where the question of the essence of the human is brought into connection with questions concerning the nature of language, philosophy, thinking and its relation to theory and praxis, ethics, and (of course) Being. This text and its broad nexus of preoccupations provide us with a starting point for considering the relation between humanity and ethics in Heidegger's thought. My concern in this chapter will be that Heidegger is unable to properly account for 'evil' and 'the inhuman' within the terms of his philosophy. I thus hope to raise a tentative concern about the political prospects of thought that renounces humanism. This concern will of course have to remain tentative in the absence of a more exhaustive survey of 'anti-humanist' positions; as I suggested in the Introduction, the views that might appropriately be called 'anti-humanist' are just as disparate as those that might be termed 'humanist'. Nonetheless, the case of Heidegger may serve to reinforce a *prima facie* wariness about such positions, and to show that the onus is very much on the 'anti-humanist' to demonstrate that their position does not threaten to imperil actual human beings and our concern for them – i.e. to demonstrate that their negation of 'humanism' need not negate what is worth preserving in that view. No

'anti-humanism' that undercuts our responsiveness to the plight of real human beings can be tenable.

But in what sense is Heidegger an 'anti-humanist'? What sense of 'humanism' does he have in mind in the title and discussions of the *Letter*? Unfortunately, these questions cannot be given a clear answer at the outset, because one of the questions Heidegger is concerned to address is precisely what sense 'humanism' still has – or whether it retains any sense at all. Indeed, the idea that the sense of this term is in question is perhaps already signalled by the scare quotes in Heidegger's title. As we shall see, Heidegger expresses ambivalence both about what importance the human being may have, and about the appropriateness of the label 'humanism' for his own later philosophy. It is therefore far from a straightforward matter to situate him in terms of a general debate between 'humanism' and 'anti-humanism', even when we adopt a capacious definition of the former term such as that canvassed in my Introduction.

Recall that I there defined 'humanism', adapting a phrase from Cora Diamond, as "the view that 'human being' is a normatively relevant notion – that is, a notion that can appropriately be used to give a reason why some creature should or should not be treated in such-and-such a way." I argued that such a definition is suitably capacious for my purposes, since it is neutral between different accounts of what the normative relevance of that notion might amount to. Furthermore, it leaves open the possibility that other notions (such as 'animal') may be normatively relevant or significant, thus obviating an identification between humanism and anthropocentrism. One way of trying to situate Heidegger in the terms of such a debate, then, would be to ask whether

he takes 'human being' to be a normatively relevant notion. In one sense, 'human being' does remain a normatively relevant notion for him: the human is given the (to Heidegger, all-important) task of becoming the 'shepherd of Being.' To this extent, it would be tempting to call Heidegger a 'humanist', albeit one of a highly idiosyncratic kind. Indeed, Heidegger himself draws attention to this possibility, and expects his interlocutor Jean Beaufret – and readers of the *Letter* more generally – to recognise its appeal: "as you have no doubt been wanting to rejoin for quite a while now – does not such thinking think precisely the *humanitas* of *homo humanus*? [...] Is this not 'humanism' in the extreme sense?" (*LH* 261)

Nevertheless, the very idiosyncrasy of Heidegger's apparent 'humanism' should give us pause. For he is equally eager to distance himself radically from other forms of humanism:

> Should we still keep the name "humanism" for a "humanism" that contradicts all previous humanism – although it in no way advocates the inhuman? And keep it just so that by sharing in the use of the name we might perhaps swim in the predominant currents, stifled in metaphysical subjectivism and submerged in oblivion of Being? Or should thinking, by means of open resistance to "humanism," risk a shock that could for the first time cause perplexity concerning the *humanitas* of *homo humanus* and its basis? (*LH* 263)

These questions are rhetorical, and it is clear what Heidegger's own answers to them would be. One of Heidegger's central concerns about 'all previous humanism' is its imbrication with what he here calls 'metaphysical subjectivism'. Humanism, on Heidegger's view, is committed to an interpretation of the human being that it takes over from the metaphysical tradition and leaves uninterrogated: the human is understood first as the *animal rationale*, and, in modernity, as the 'subject'. This

conception of the human has, in turn, deep connections to the rise in late modernity of what Heidegger calls 'technology', and the ever-increasing objectification of beings that characterises its development. Heidegger's later writings constitute a sustained attempt to bring the 'danger' of technology to light, and thus stand in 'open resistance' to the tacitly metaphysical outlook of other forms of humanism.

Beyond Heidegger's own doubts about the aptness of the label 'humanism' for his later thinking, we may also wonder further whether Heidegger takes the notion of the 'human being' to be normatively significant in anything like the sense indicated in the second part of my capacious definition. In other words, it is unclear whether Heidegger would take the mere fact that a creature is a human being to give us a reason to treat – or refrain from treating – that creature in such-and-such a way. In general, there is a striking absence of specific moral prescriptions in his philosophical work, for reasons that I will try to articulate later. Thus it is unclear what stand Heidegger's philosophy takes on questions about what we owe to each other, or what is owed to us, simply in virtue of the kind of entities we are – if it takes a stand on such questions at all. Though Heidegger claims that he 'in no way advocates the inhuman', it would not follow that his philosophy gives us any robust grounds for resistance to it. Furthermore, there may be reasons to worry that the familiar ethical distinction between the human(e) and the inhuman(e) simply cuts across the version of the distinction Heidegger himself is concerned to draw in the *Letter*.

The foundation for this worry will become clearer in due course. First, a word about the context and concerns of the *Letter* itself. The proximate occasion for Heidegger's

(1946) *Letter* was a series of questions from Jean Beaufret, some of which Heidegger quotes directly and others of which he only alludes to. These questions, which principally concern the relationship between humanism, existentialism, ontology, and ethics, were in turn prompted by Jean-Paul Sartre's (1945) public lecture *Existentialism is a Humanism*. Sartre's text, which was published the following year, paraphrased a Heideggerian formula from *Being and Time* – that for 'Man' "existence precedes essence" – and claimed it as the fundamental tenet of existentialism, arguing (as Sartre's title suggests) for a humanistic interpretation of that doctrine. In his *Letter*, which briefly addresses Sartre's claim and censures it as a misunderstanding, Heidegger is thus in dialogue with Beaufret, with Sartre, and perhaps most extensively, with his own earlier work. He argues for an interpretation of *Being and Time* that not only sets itself 'against humanism', but also begins to clear the way for a conception of the essence of the human that surpasses the possibilities of humanistic thought:

> [T]he highest determinations of the essence of the human being in humanism still do not realize the proper dignity of the human being. To that extent the thinking in *Being and Time* is against humanism. But this opposition does not mean that such thinking aligns itself against the humane and advocates the inhuman, that it promotes the inhumane and deprecates the dignity of the human being. Humanism is opposed because it does not set the *humanitas* of the human being high enough. (*LH* 251)

In opposing Sartre's humanist appropriation of *Being and Time*, Heidegger is thus concerned to show that it is possible (indeed, necessary) to think 'against' humanism, understood as a fundamentally and unacceptably metaphysical doctrine. But he nonetheless claims to recognise the "essential worth" of the human being (*LH* 252), and, further, to show that a rejection of humanism need not mean 'advocacy' and 'promotion' of the inhuman. This latter task is in fact a recurring concern of the *Letter*,

and Heidegger spends a substantial number of its pages attempting to dispel certain possible misreadings of his talk of 'speaking against' – not just 'against' humanism, but against logic and values too (*LH* 263ff.). Indeed, as can be seen in the passage cited above, Heidegger goes further, suggesting that it is in fact humanism *itself* that represents the danger of deprecating the human, in failing to 'set the *humanitas* of the human being high enough'. Heidegger thus seeks to provide us with an account of the essence of the human being that succeeds in 'realizing its proper dignity'.

The task of providing a defence against accusations of 'advocating the inhuman' arises naturally enough from a discourse that seeks to question humanism. This is because thinking 'against humanism' may all too easily be construed as thinking against the 'values' that humanism represents, chief among them the 'essential worth' of the human being. Heidegger himself acknowledges this potential objection, writing that "because we are speaking against 'humanism' people fear a defense of the inhuman and a glorification of barbaric brutality" (*LH* 263). The underlying thought seems to be that the opposite of humanism is a kind of barbarism, so that opposition to the former becomes a tacit endorsement of the latter. We could view Heidegger's efforts to answer this charge, then, as responsive to an intrinsic demand of the subject matter. However, there is also a striking and significant extrinsic cause for concern here – or rather, a cause for concern whose in- or extrinsicality to the subject matter at hand remains an open question. I refer here to Heidegger's (1933-4) involvement with National Socialism, and his subsequent failure to unequivocally denounce the atrocities committed by the Nazis in the 30s and 40s, most notably and disquietingly the holocaust. Heidegger's proximity to these events makes a critical inquiry into his

opposition to humanism all the more urgent, and it is incumbent on us to ask what relation there might be between his apparent unwillingness or inability to speak about the evil of the holocaust, and his own philosophical claims about evil and the inhuman.

Granting that Heidegger does not go so far as to 'advocate the inhuman', it still remains to be determined whether or not his position is politically and ethically inadequate, for it is surely the case that a philosophy could avoid such advocacy while nonetheless failing to provide – or even obscuring or impoverishing – the resources needed to properly acknowledge, criticise, or even think through the inhuman. We must therefore ask how Heidegger understands 'the inhuman' and 'evil' in relation to his original understanding of the essence of the human, and consider whether this understanding can be adequate as a response to and an account of the evil, horror, and inhumanity of the events of the 20th century – or whether, on the contrary, Heidegger's apparent disavowal of 'ethics' is implicated in his failure to adequately acknowledge those events.

In Section I, I briefly trace some key developments in Heidegger's thought regarding the human, so that the claims about the 'essence of the human' in the *Letter* may be seen in the right light. I argue that the common impression of anthropocentrism in *Being and Time* results from its extant divisions' focus on the 'existential analytic' of Dasein, the entity which Heidegger seemingly identifies with 'man' and with 'we ourselves'. However, this impression is misleading, because that analysis is intended as a 'preparatory' way into the question of Being, and is subordinate to it. Shifts in vocabulary and methodology in Heidegger's later work can be understood in part as a

response to this charge of anthropocentrism. The later Heidegger retains an essential role for the human being, who is appropriated, 'needed and used' by Being, and whose essence is attained in the preservation of Being's truth. But this happens "only to the honour of Being," not "for the sake of the human being" (*LH* 251). The human is radically decentred, yet remains essential.

In Section II, I elucidate some of the central themes of the *Letter* itself, and explain Heidegger's rejection of humanism and the notion of 'values'. In brief, humanism misunderstands the human essence, thus occluding our relationship to Being; the critique of 'value' does not target normativity as such, but rather the vocabulary of valuation, which objectifies beings, depriving them of their 'dignity' by making their worth dependent on subjects. This discussion will allow me to shed light on Heidegger's more positive claims concerning the essence and proper dignity of the human being. Finally, I introduce Heidegger's remarks on evil in the *Letter* in order to illuminate his normative aims and his conception of 'the inhuman'.

These remarks are brief and cryptic, so I turn in Section III to the fuller treatment of evil in the third of Heidegger's *Country Path Conversations*.[1] I argue that the account developed there involves a problematic displacement of human responsibility for evil, which is reinterpreted as merely symptomatic of a broader movement in the history of Being (namely a development in late modernity that Heidegger calls the abandonment of beings by Being). This diagnosis of evil is coupled with a disavowal of moral critique and of the appropriateness of guilt, as well as a de-racialized but nonetheless suspect

[1] I am grateful to Mark Wrathall for pointing me in the direction of this text as a source for Heidegger's views on evil.

centring of the German people. The result is an unacceptable quietism about evil in the 'ordinary' sense, which perhaps goes some way toward explaining Heidegger's conspicuous silence regarding the horrors of the Nazi regime. Although to some extent separable from other elements of his later philosophy, these ideas are nonetheless bound up with his conception of the abandonment of Being, and by extension with the central theme of a connection between technology and nihilism. The considerations developed here, then, cast a troubling light on key elements of Heidegger's later philosophy, and should occasion unease and caution regarding his thinking 'against' humanism.

I. The Question of the Human in Heidegger's Philosophy

It is not surprising that many readers have taken Heidegger's *Being and Time* to be a work whose chief concern is to provide an analysis of specifically human existence. Heidegger announces in the brief preface that his "aim in the following treatise is to work out the question of the meaning of *Being* and to do so concretely" (*BT* 19/H1). What interests Heidegger is evidently the question of Being as such, rather than questions concerning the Being of this or that particular entity. Nonetheless, it quickly becomes clear that there is an entity that occupies a privileged position in Heidegger's inquiry into the question of Being, namely 'Dasein'. This is because Dasein is the entity who has "the very asking of this question" as a "mode of *Being*" (*BT* 27/H7) – that is to say, Dasein is the entity for whom questioning and inquiring are a possibility. Moreover, Dasein is actually defined by its special relationship to Being: "*Understanding of Being is itself a definite characteristic of Dasein's Being.* Dasein is ontically distinctive in that it *is* ontological" (*BT* 32/H12). Naturally, then, Division I of this work of

fundamental ontology takes the form of a 'Preparatory Fundamental Analysis of Dasein'. Moreover, while the German term 'Dasein' usually means 'existence' or actuality in a quite general sense (such that there would be no oddity in talking about the 'Dasein' of animals, plants, stones, etc.), Heidegger explicitly restricts its usage. 'Dasein' refers to the kind of entity that 'we ourselves' are: as Heidegger puts it in the very first sentence of the first chapter of Division I, "We are ourselves the entities to be analysed" (*BT* 67/H41).[2] In case the referent of this 'we' should appear indeterminate, we might recall that earlier still, Heidegger writes that "man himself" is "the entity we denote by the term '*Dasein*'" (*BT* 32/H11). Furthermore, the phenomenological analyses that follow – of equipment, being-with, understanding and mood, anxiety, death etc. – all take a specifically *human* mode of Being as their point of departure. The impression that *Being and Time* is in large part a contribution to philosophical anthropology – albeit one that is ultimately oriented toward and subordinated to broader ontological questions – is by no means ungrounded in the text.

Of course, there are different ways of taking Heidegger's claims, and further complexities to consider. We saw in Chapter One that Kant took his 'formula of humanity' to apply not only to human beings but to every rational being. As we shall see, Heidegger will want to strongly resist any identification between 'Dasein' and the *animal rationale*. But one might think that an analogous generalization to that made by

[2] Heidegger makes a corresponding restriction regarding the use of the term 'existence' ('Existenz'): "the term 'existence', as a designation of Being, will be allotted *solely* to Dasein"(*BT* 68/H42, emphasis added). In Heidegger's terminology, *only* Dasein can be properly said to 'exist'. This restriction is motivated in part by Heidegger's desire to distinguish terminologically between Dasein's Being and that of other entities, for instance the 'present-at-hand'. Indeed, in the very paragraph in which Heidegger restricts the usage of 'existence', he distances it from the Latin '*existentia*', and introduces the term '*presence-at-hand*' as an 'Interpretative' gloss on the latter.

Kant might nonetheless be possible. If Heidegger explains 'Dasein' in terms of its understanding of Being and the other *existentialia*, we might wonder about the extension of this term. Heidegger himself seems to resist the idea that animals (and indeed God) could have the kind of Being that Dasein has.[3] But are *only* human beings 'Dasein'? Are *all* human beings 'Dasein'? There are also interesting questions regarding how we are to understand Heidegger's 'we' and its function in determining the semantics of 'Dasein'.[4]

While I shall bracket the majority of these questions in what follows for reasons of space, I will make two interrelated points here that will provide crucial background for the remarks Heidegger makes about the human essence in the *Letter*. Firstly, in his writings in the 30s, Heidegger himself increasingly came to question his own invocation of the first person plural. In a section of the *Contributions* entitled '*Philosophy* (On the question: Who are we?)', Heidegger embarks on a sustained interrogation of that question, and of various potential answers to it: 'those here and now', '"the" human being as such', 'this particular people' etc. (Ga65: 39-44, §19). Furthermore, he refers the reader to his 1934 lecture course on *Logic as the Question Concerning the Essence of Language*, in which he argues for a shift from the "What-question [that] conceives the

[3] See especially Ga49: 29, and the discussion of how the animal is "world poor" in Ga29/30. See further the debate concerning Heidegger's understanding of the relationship between 'world' and non-human animals in Derrida (1989), Chapter VI; Mulhall (2008b), Chapter 6; and, most extensively, Cykowski (2021). While the question of Heidegger's treatment of non-human animals is obviously relevant to my broader concerns about anthropocentrism, I shall put it aside here. As will become clear, Heidegger has his own criticisms of anthropocentrism, though strikingly he is much more concerned to reorient us towards *Being*, rather than towards any specific set of entities (such as animals and the environment).

[4] Wayne Martin endorses an 'exemplarist' reading, according to which Heidegger and his reader are *exemplars* of the type of entity that Dasein is: "'Dasein" refers to Heidegger and to me and to anything else that shares our mode of being, our ontological structure"; see Martin (2013: 118).

human being as a present-at-hand thing" – i.e. the question 'What is the human being?' – to the 'Who-question': "*Who* is the human being?" (Ga38: 35) This shift is already suggested in *Being and Time*'s claims that "because Dasein has *in each case mineness* [*Jemeinigkeit*], one must always use a *personal* pronoun when one addresses it: 'I am', 'you are'" (68/H42), and that "any entity is either a "*who*" (existence) or a "*what*" (presence-at-hand in the broadest sense)" (71/H45). Dasein, of course, is not merely a *what*, but a *who*: an 'I' or a 'you' – or indeed a 'we'. But Heidegger also takes the consideration of the question 'who are we?' in a different direction. This leads me to my second point: Heidegger increasingly resists a straightforward identification between human beings as such and Dasein, even while affirming an intimate relationship between the two terms: "Only in the domain of the Being of the human, even if not in every humanity *and perhaps not yet even in any of them*, does what is now called 'Dasein' show its essence" (Ga49: 27, emphasis added).

Heidegger makes this claim in the context of a highly illuminating discussion that he titles an 'Elucidation of the Concepts of Existence and Da-sein (Elucidations of "Being and Time")'.[5] The discussion is framed primarily as a clarification of these two notions, which, as I suggested above, are absolutely central to *Being and Time*. But given what I said in the opening of this section, Heidegger's efforts to problematize the relation

[5] It is unclear whether Heidegger provides what he takes to be a plausible reading of his earlier text, or an attempt at an immanent critique of it. Heidegger later acknowledges that some of his formulations in *Being and Time* are liable to encourage certain misinterpretations; see in particular his even stronger statement in the *Letter*: "In the meantime I have learned to see that these very terms were bound to lead immediately and inevitably into error" (*LH* 271). Thus, the 'elucidation' may be best understood as a 'strong' reading of the earlier work – but one that can be taken to arise from some of the central ideas of that text, and which Heidegger could plausibly claim better captures its spirit: "we are now able to say certain things about the necessity of being and time more clearly than was possible in the first attempt of the treatise 'Being and Time'" (Ga49: 27).

between humanity and Dasein are likely to come as a surprise. The terms are often treated as all but interchangeable. 'Dasein' (or 'Da-sein') is presented in some of Heidegger's later work as more of a possible *achievement* of humanity than as a determination of its basic way of Being. This way of using the term seems, at least on the face of it, to be in tension with Heidegger's usage in *Being and Time*. Why, then, does he present his 'Elucidation' as a clarification of the notions developed in that text, or as a corrective against what he now takes to be pervasive forms of misinterpretation?

To get clearer on this question, let us consider in outline the interpretations from which Heidegger recoiled. There are three interpretative claims that Heidegger seems to be attempting to distance himself from: that *Being and Time* was a work of philosophical anthropology; that the views expressed in it lead to an extreme subjectivism; and that Heidegger's position was a form of idealism. In fact, these interpretative claims are intimately related, and arguably mutually reinforcing. If *Being and Time* is a work of philosophical anthropology, it seeks to understand Being by understanding the human being (Dasein). Moreover, Heidegger seems to make the former dependent on the latter: "only as long as Dasein is (that is, only as long as an understanding of Being is ontically possible), 'is there' Being"; "Being (not entities) is dependent upon the understanding of Being" (255/H212). Even 'truth' seems to become dependent on Dasein and therefore on the human being. As Heidegger emphatically puts it: "*'There is' truth only in so far as Dasein i s and so long as Dasein i s*" (269/H226). The distance between these claims and full-blown idealism will seem, to many, very small, and the sense of proximity to idealism is hardly allayed by Heidegger's claim that "If what the term 'idealism' says, amounts to the understanding that Being can never be explained

by entities but is already that which is 'transcendental' for every entity, then idealism affords the only correct possibility for a philosophical problematic" (251/H208).[6] It is thus not entirely surprising that one of Heidegger's contemporaries, Nicolai Hartmann, felt able to claim:

> The consequence of [Heidegger's] approach is that everything that is, is from the start understood as relative to the human. It is in each case his own. All further determinations then result from this relativization to the "I" of the human: the world, in which I am, is "in each case mine," and could very well be, for each, another world; in the same way truth is "in each case mine." [...] The real fallacy in the approach is that Being and the understanding of Being are brought too close together; Being and the givenness of Being are virtually conflated.[7]

Now, much of this interpretation is not just mistaken, but, on even a cursory reading of *Being and Time*, clearly uncharitable: the gloss on '*Jemeinigkeit*', the lack of acknowledgement that for Heidegger the world is basically a public and social phenomenon, and the particular emphasis put on the 'I' are especially striking. Nonetheless, the idea of a relativization to the human is certainly encouraged by Heidegger's apparent identification between Dasein and the human and by his remarks on the dependence of Being on Dasein and its understanding of Being.

Even so, there is much in the early work that, as Heidegger later points out, is recalcitrant to such an interpretation. Indeed, it is recalcitrant in ways that raise questions about the straightforward identification between Dasein and the human. For already in *Being and Time*, Heidegger very explicitly distances himself from philosophical

[6] See Moore (2012), Chapter 18, §7 for further discussion of idealism in Heidegger's work, including a discussion of the passages just quoted.
[7] This passage is cited and discussed by Heidegger in Ga49: 50-1; all occurrences of 'Being' have been capitalized.

anthropology, and regards the latter as (so far) philosophically ungrounded. In the Introduction, he emphasises that "the analytic of Dasein remains *wholly* oriented towards the guiding task of working out the question of Being. [...] It cannot attempt to provide a complete ontology of Dasein, which assuredly must be constructed if anything like a 'philosophical' anthropology is to have a philosophically adequate basis" (38/H17, emphasis added). He later reiterates that "The existential analytic of Dasein comes *before* any psychology or anthropology, and certainly before any biology" (71/H45), and draws attention to his intentional avoidance of the term 'man' as a designation of "those entities which we ourselves are" (72/H46).[8]

Nonetheless, the fact remains that, even if he subordinates this question to the orienting question of Being, Heidegger appears to be offering an account of the basic and necessary structures of Dasein's existence, and therefore of the Being of the human being. How, then, can we make sense of Heidegger's conception of Dasein in a way that distinguishes his project from philosophical anthropology? Heidegger's answer in the 1941 lecture course is worth quoting at length:

> What is meant is not "*Da*sein" in the sense of the presence of a thing or of the human that is here and there and "*da*"; rather, what is being thought is "Da-*sein*," that the clearing for Being in general essences and is. The word "*da*" is not the name for a location, whether it be in particular or in general; rather, the "*Da*" names the clearing within which a spatio-temporal ordering of positions and in general space and time in the ordinary sense can be posited.

[8] See Mulhall (2001: 215-8), for discussion of Heidegger's distinction between his existential analytic and the disciplines of anthropology, psychology, and biology. As Mulhall points out, these disciplines seek to study in isolation the three 'components' which have traditionally been understood to make up the Being of the human being: spirit, soul, and body. Heidegger is rejecting the idea that our Being can be correctly understood as a compound of these three elements, taken individually. This rejection goes hand in hand with his rejection of the traditional designation of the human being as *animal rationale* – as a living creature endowed with reason.

> Da-sein, understood in this way, is called "human," not because it is grounded in the being of the human and is something "human," something measured according to the human, but because, at the most, "the" human is specifically assigned to Da-sein. "The human" – that is, a historical humanity {*Geschichtlichkeit*} [...] (from out of the persistent steadfastness taken over from the destinal {*geschicklichen*} experience of Being). But not every historical humanity is specifically assigned to the persistent steadfastness of Dasein; as a result of the forgetting of Being, which is to be thought in terms of the destiny of Being, *there has not yet been such a humanity in history up to this point*; for, all of Western History, and the modern history of the world in general, is grounded in metaphysics. But a humanity to come is delivered over to Da-sein. (Ga49: 47-8, final emphasis added)

There are certainly some deep obscurities here, most notably Heidegger's highly idiosyncratic use of 'essences' as a verb and the unfamiliar notions of the clearing for Being and the destinal experience of Being. It is also far from clear what might be involved in the 'persistent steadfastness of Dasein'. While I cannot hope to give a full account of these ideas here, they nonetheless provide an initial indication of the basic contours of Heidegger's later conception of the human and of Dasein. First, while the human being and Dasein remain intimately related, the relation is not understood as one of identity. Rather, 'Da-sein' is something to which the human may or may not come to be 'assigned'. Moreover, this 'assignment' is related to a 'destinal experience of Being'. Some connotations of this phrase are worth drawing out. First of all, the phrase might suggest that (some future) humanity is *destined* for or fated to a certain experience of Being. This in turn suggests both a fundamentally historical dimension to this experience and a certain passivity or receptivity on the part of the human being, which is further reinforced by the idea that humanity might be 'delivered over' to Dasein. Indeed, the idea that the history of Being is somehow prior to (and enabling of) the deeds of human beings is a recurring one in Heidegger's later work, which we shall encounter again in Section II.

A second point is that the notion of an experience *of Being*, and indeed of a forgetting *of Being*, suggests that Being has a certain independence or otherness. This complicates the suggestion in *Being and Time* that Being is 'dependent' on Dasein – or rather, it suggests a very different interpretation of that claim. Elsewhere in his later work, Heidegger talks of a "dislodging" of the human being, but cautions against the idea that this is something that humans could achieve of their own accord: "If human beings must carry out this dislodging by relying on their own resources, then is not the presumptuousness of giving the measure even *greater* than it is when the human being is simply put forth as the measure?" (Ga65: 22, §7) There is a noteworthy parallel between the idea of the human being 'giving the measure' expressed here and the similar language of Dasein being (misunderstood as) 'something measured according to the human' in the passage quoted above. The passage in the *Contributions* goes on to suggest that the 'dislodging' of the human is something that could "befall" them, an event which is immediately related back to the history of Being. So throughout both discussions, *Being*, and not the human, is understood as the source and measure of the understanding of Dasein.

What, then, of *Being and Time*'s claim that 'Being is dependent upon the understanding of Being', and so on Dasein? One clue to understanding Heidegger's later transformation of this claim is the notion of 'assignment'. This term is suggestive of a task for 'Dasein' and the human being. Indeed, we are told that "Being, in its uniqueness, *needs* Da-sein and also *needs* what is therein both grounded and grounding, namely, the human being. Truth does not essentially occur otherwise" (Ga65: 251, §194, emphasis added). Here we can hear echoes of the claims from *Being and Time* that

both Being and truth (understood as unconcealment) are dependent on Dasein. But that dependency is now given a quite specific interpretation. The term 'needs' translates '*braucht*', which could also be translated as 'uses'. So one connotation is that Heidegger envisions the humanity of the future as, in a certain sense, in the service of Being. This notion, finally, brings us to the ideas about the essence of the human being expressed in the *Letter on 'Humanism'*. In a striking passage, Heidegger tells us: "The human being is not the lord of beings. The human being is the shepherd of Being. [...] They gain the essential poverty of the shepherd, whose dignity consists in being called by Being itself into the preservation of Being's truth" (*LH* 260). This passage confirms the idea that we are somehow beholden to, 'called by' Being. And we are called to a specific task: the preservation of Being's truth. It is in this sense that "the essence of the human being is essential for the truth of Being" (*LH* 263). The passage also points to an inadequate conception of the human: as the *lord* of beings. This misinterpretation in turn indicates another central element of Heidegger's later thought: the idea that the essential characteristics of our modern, technological age are the ever-increasing *mastery*, manipulation, ordering, calculation, and objectification of beings – perhaps even ourselves – and the oblivion of Being.

II. The *Letter*: Heidegger Contra 'Humanism'?

What do these considerations have to do with 'humanism'? In this section, I will provide a close reading of some of the central contentions of Heidegger's *Letter* in order to shed light on this question, and so on Heidegger's later conception of the human essence and the proper dignity of the human being. First, I will explain Heidegger's critique of 'all previous humanism', and how this relates to his new understanding of

the human essence: 'ek-sistence'. Second, I will explain Heidegger's critique of 'values', and show how his discussion of this notion sheds light on his idiosyncratic conception of dignity (including human dignity). Third, I will turn to Heidegger's discussion of evil and inhumanity in the *Letter*, to evaluate whether his anti-humanist position can provide an adequate account of – and opposition to – the inhuman.

A. The Essence and Ek-sistence of Humanity

Heidegger's later conception of *humanitas*, the human essence, can perhaps be clarified by way of a contrast with his target in the *Letter*, which he identifies as "all previous humanism" (*LH* 263). What, in Heidegger's view, unites all previous forms of humanism, so as to allow him to group this apparently disparate set of historical positions together? How does he distance his new conception of the human essence from this earlier, supposedly unified tradition?

An answer to these questions can be found in the connection, prominent in the *Letter*, that Heidegger draws between humanism and the metaphysical tradition. His argument for the sweeping claim that "every humanism is grounded in a metaphysics or is itself made to be the ground of one" (*LH* 245) is in some respects very simple. At this stage in his development, Heidegger defines metaphysics in terms of its failure to raise the question of Being: it "does not think Being as such, does not think the difference between Being and beings" (*LH* 246). For that reason, metaphysics does not "ask in what way the essence of the human being belongs to the truth of Being" (*LH* 246); it does not ask about the relationship between Being and the human being, because it does not ask about Being at all. But for Heidegger, our "*humanitas*, the essence of the

human being [...] lies in ek-sistence" (*LH* 262), which he glosses elsewhere as "standing in the clearing of Being" (*LH* 247). The essence of the human being can thus only be understood in terms of its relationship to Being. Humanistic understandings of the human being never so much as raise the question of Being, and so fail to locate the human being in relation to it. In this sense, they are one and all 'metaphysical'.

But why does Heidegger view ek-sistence as our essence? Because ek-sistence is a "way of being [...] proper *only* to the human being" (*LH* 247, emphasis added). Ek-sistence is the way of being that radically distinguishes us from other kinds of entity (including, so Heidegger claims, other animals), thus recalling his earlier conception of Dasein. Indeed, Heidegger writes that "the human being ek-sistingly sustains" Da-sein, suggesting that he continues to associate the two concepts (*LH* 251; see also 249). In the *Contributions*, the language Heidegger uses to talk about 'Da-sein' echoes what he says in the *Letter* about ek-sistence: the human being is the entity "to whom alone Da-*sein* is *proper*", and Da-sein in turn is understood as "the mode of *being* that is distinctive of humans *in their possibility*." This final point, however, again draws attention to the shift in Heidegger's conception of Dasein – now, Da-sein – that I discussed in the previous section. For he goes on to ask, "In which possibility?" and to answer: "In the highest one of humans, namely the possibility of grounding and preserving truth itself" (Ga65: 237, §176).

The fact that Heidegger identifies this as our *highest* possibility suggests that, though this possibility is the one most proper to us, it is not one that we always actualise. It would make little sense to talk of a way of being as the highest possibility if there were not

various lower possibilities with which it could be contrasted. In the *Contributions*, Heidegger elaborates by distinguishing between Da-sein (literally, being-there) and being-away (*Weg-sein*): "Da-*sein*: withstanding the openness of self-concealing. *Being*-away: pressing on with *the closedness of the mystery and of Being*; forgottenness of Being. And this happens in *being-away* according to this sense: *to be infatuated with things, smitten with them, lost in them*" (Ga65: 238, §177).[9] This formulation presages claims Heidegger will make later, both in the *Letter* and in his essay on technology. For the danger of technology is to be understood in terms of the way it closes us off from other possible understandings of Being, from what the *Contributions* calls the 'mystery': "The actual threat has already affected man in his essence. The rule of Enframing threatens man with the possibility that it could be denied to him to enter into a more original revealing and hence to experience the call of a more primal truth" (*QCT* 28). In the same paragraph, Heidegger appears to allude to his claim from the *Letter* that humanism and metaphysics culminate in the view that man is the 'lord of beings': "man, precisely as the one so threatened, exalts himself to the posture of lord of the earth. In this way the impression comes to prevail that everything man encounters exists only insofar as it is his construct. [...] It seems as though man everywhere and always encounters only himself." But on the contrary, insofar as we remain deaf to the 'call' of Being, "*precisely nowhere does man today any longer encounter himself, i.e., his essence*" (*QCT* 27). In '*Weg-sein*', being-away, human beings remain alienated from their essence. They remain unable to preserve the truth of Being.

[9] Heidegger goes on to tacitly acknowledge this new concept's proximity to *Being and Time*'s notion of inauthenticity, but at the same time expresses his dissatisfaction with his earlier terminology.

This is why, from the very opening of the *Contributions*, Heidegger identifies the task of philosophy as "no longer to be 'about' something, to present something objective", but rather to bring about "an essential transformation of the human being: from 'rational animal' (*animal rationale*) to Da-sein" (Ga65: 5 [Unnumbered Preface]). This suggests that the basic point of his later work is, in a broad sense, ethical. It is not (just) about describing what the human essence is, but about transforming the human being. Passages like this one also seem to me to undermine – or at least to complicate – Thomas Sheehan's claim that "human beings are the *Da* [of Da-sein] not occasionally or by their own choice, but *of necessity*. We cannot *not* be the open."[10] While such a claim is certainly encouraged by the manner in which Heidegger talks about Dasein in *Being and Time*, it shows insufficient sensitivity to the shifts in his terminology. It is surely significant that Heidegger explicitly understands '*Weg-sein*', the apparent opposite of Da-sein, in terms of the metaphor of 'closedness'. Thus, as Richard Polt has put it, "what was presented in *Being and Time* as a phenomenology of Dasein – understood as the human way of being – should instead be conceived as projecting a new possibility for humanity."[11] Da-sein is still viewed as a uniquely or distinctively human way of being – indeed, as the highest human possibility – but it is no longer viewed as the way humans are in general or necessarily. There is one complication, however, namely that we are never *completely* closed off from Being. The passage from the technology essay continues: under the spell of 'enframing', man "fails in every way to hear in what respect he ek-sists, from out of his essence, in the realm of an exhortation or address, and thus *can never* encounter only himself" (*QCT* 27). Even when we have forgotten Being (or as he sometimes puts it, been abandoned by Being), the possibility of an

[10] Sheehan (2001: 194).
[11] Polt (2020: 55).

openness to Being, to its 'address' or 'call', remains. In a sense, then, Being and the human being remain indissociable. Nonetheless, it seems clear that Heidegger wants to reserve the term 'Da-sein' for a specific way of being: one in which we do not fail to heed our essence and ek-sistence; one in which we preserve the truth of Being.

Before we return to the question of humanism, allow me to briefly consider two critical questions that might be thought to arise naturally from the account I have given so far. Indeed, both questions arise from the highly abstract character of what Heidegger has said so far. First, what does the 'essential transformation' of the human being that Heidegger calls for actually consist in? Second, how are we to understand Heidegger's use of the word 'Being'? In particular, how "metaphysically inflationary" is Heidegger's conception of Being?[12] There are severe difficulties involved in answering the first question. It is hard to give a concrete account of what the transformation of our essence might amount to, not least because Heidegger himself sometimes claimed to be at a loss. In his final interview with *Der Spiegel*, he claims that his later writings can at best awaken a receptive mood in us, or begin to prepare us for another way of thinking. What he cannot give, however, is concrete moral direction.[13] Indeed, he sometimes suggests that the transformation his philosophy is supposed to engender is not to be understood in moral or narrowly ethical terms at all: "In philosophical knowledge [...] the very first step sets in motion a transformation of the one who understands, and this not in the moral-'existentiell' sense, but rather with respect to Da-sein" (Ga65: 13, §5; see also 238, §177). One explanation for Heidegger's reticence here is the sheer variety of concrete ways of being – what he would earlier have called

[12] I borrow the quoted phrase from Golob (2022: 509).
[13] See e.g. *OAG* 107-110, 114-5.

'existentiell' possibilities of Dasein – that might count as a way of "sheltering the truth [of Being] in some fashion or other": we can do this, he suggests, "thoughtfully or poetically, or by building, leading, sacrificing, suffering, rejoicing" (Ga65: 238, §177). This strongly suggests that whether a 'transformation' has taken place is not so much a question of the content of our deeds, but of the manner in which they are (or fail to be) responsive to Being. Heidegger writes in the *Letter* that ek-sistence is a way of being "in which the essence of the human being preserves the source that determines him", i.e. Being itself (*LH* 247). His language here suggests both an active and a passive or receptive moment: the ek-sisting human being both (actively) *preserves* or *shelters* and is (passively) *determined by* Being.

But this makes the second critical question all the more pressing. How are we to understand 'Being' in all of these claims? As Sheehan and Golob have pointed out, there is surely a danger of reifying 'Being', and many of Heidegger's formulations appear to succumb to this danger by describing Being as an active and quasi-agential force – what Sheehan sardonically names 'Big Being' – while downplaying the agency of human beings.[14] In a representative passage from the *Letter*, Heidegger stresses that "human beings do not decide whether and how beings appear, whether and how God and the gods or history and nature come forward into the clearing of Being, come to presence and depart. The advent of beings lies in the destiny of Being" (*LH* 252). It is the 'destiny' of Being that appears to determine what kind of understanding of Being

[14] See Sheehan (2001: 189). Sheehan, provocatively and in my view too hastily, cites this tendency as a reason for abandoning the word 'Being' as a label for the central subject matter of Heidegger's philosophy. See also Golob (2017: 630-3) for illuminating discussion of these issues.

human beings can have, and thus to determine the ways in which beings, entities, can appear to us.

I shall return to the questions this raises concerning human agency – and indeed responsibility – in §III. For now, however, it will be sufficient to note that such passages do not necessarily settle the question of how 'metaphysically inflationary' Heidegger's conception of Being needs to be. While they do lend themselves to a hypostatisation of Being, they do not rule out a more deflationary reading.[15] Part of the point could simply be that human beings cannot in any straightforward sense choose their understanding of Being. They are, in the language of *Being and Time*, thrown into such an understanding. The question will then be whether modern human beings remain fixed, stuck in the conception of beings into which they find themselves thrown – whether this conception takes beings to be 'objects' facing a metaphysical subject, or to be 'standing-reserve', mere resources for human ends – or become receptive or open to other understandings of Being, and thus of other ways in which beings can be. Rather than remaining within the confines of a one-sided interpretation of entities, an attitude of responsive thought might unable us to "guard the truth of Being, in order that beings might appear in the light of Being as the beings they are" (*LH* 252).

[15] There will also be intermediate positions. One interesting suggestion in this regard is Miguel de Beistegui's view that Heidegger understands Being in his later work as a 'pure event', an event without a subject; see de Beistegui (2003: 221-6). *Es gibt*, 'there is' (literally 'it gives'), would thus be understood on the model of *es regnet*, 'it is raining'. In both cases it would be a misunderstanding to construe the 'it' as referring to a subject or entity, so when Heidegger tells us that Being is itself the 'It' of *es gibt* (*LH* 252), we need not take him to be claiming that Being is a mysterious, active subject doing the 'giving'. Furthermore, this conception coheres well with the later Heidegger's complex but pervasive use of the term '*Ereignis*' ('event') and its cognates in relation to the truth of Being. Nonetheless, a great deal more would need to be said to fully unpack and evaluate this suggestion and the ontic metaphor on which it relies, and I do not have the space to do so here.

For my purposes, we can bracket the question of whether a more inflationary or deflationary reading makes the best sense of Heidegger's claims. The important point is that Heidegger's new conception of the human being understands humanity primarily in terms of its (possible) relationship to Being. As he emphasises in the *Letter*, "the point is that in the determination of the humanity of the human being as ek-sistence what is essential is not the human being but Being" (*LH* 254). This point also allows us to see what Heidegger has in mind when he talks about the *in*human: for a human to be "inhuman" is for them to be "outside their essence" (*LH* 244). Our essence is ek-sistence, which has come to mean standing in the clearing of Being and preserving its truth. This suggests that in Heidegger's sense, 'inhumanity' is to be understood ontologically. To be 'inhuman', Heidegger implies, is to "swim in the predominant currents, stifled in metaphysical subjectivism and submerged in oblivion of Being" (*LH* 263). It is thus metaphysical humanism that perpetuates the inhuman in Heidegger's sense. Inhumanity, like humanity, must be characterised in terms of the human being's relationship to Being.

If we understand the inhuman in *this* sense, it is hardly a surprise that Heidegger claims he in no way advocates or defends the inhuman (*LH* 251, 265). But this does not yet settle the question of the relationship between this sense of the inhuman and its more familiar ontic sense, which relates to inhumane actions, deeds, policies, and institutions. Thus, although I have indicated a broad sense in which Heidegger's later work is oriented toward ethical transformation, and thus has a pronounced normative impetus, it is still unclear how such a transformation relates to what we ordinarily understand by 'ethics' and ethical action. This is an important issue, and I shall return

to it. First, however, I want to consider a more basic objection, namely that some of his claims in the *Letter* might appear to undermine any normative interpretation of his claims about the human essence. In particular, Heidegger subjects the idea of 'values' to severe criticism. Given the conceptual connections that are often drawn between values, the 'ought', and normativity, one could be forgiven for thinking that Heidegger's rejection of 'values' seems to amount to a rejection of normativity as such.

In the next part of this section, I will try to answer this objection by giving an interpretation of what Heidegger's complaints 'against' values amount to, and showing that those complaints are compatible with what I have identified as the broadly ethical intent of his later writings. Indeed, by making sense of Heidegger's rejection of the notion of 'values', we can shed light on one of the central aspects of his critique of humanism: his rejection of the conception of the human being as a 'subject'. The notion of the 'subject', on Heidegger's view, is a particular historical development or interpretation of the humanistic view of the human being as the rational animal (*LH* 245-7; cf. 254). It will turn out that the very vocabulary of 'values' is, on Heidegger's account, intimately bound up with the interpretation of the human being as subject and the attendant interpretation of beings in general as *objects*. In other words, talk of 'values' will turn out to be covertly implicated in the metaphysical conception of the world that becomes dominant in modernity, and is massively intensified in what Heidegger views as the present age of technology.

B. Thinking Against 'Values'

Heidegger is acutely aware that a blanket critique of 'values' will be construed all too

easily either as a rejection of everything that we judge to be valuable, or as amounting to the claim that there are no values, and thus that nothing at all is valuable. Nevertheless, he claims that such a conclusion would be based on a misunderstanding – one that exactly parallels the worry that the rejection of humanism amounts to an affirmation of inhumanity. Indeed, he suggests these "misinterpretations [...] betray the same structure and the same foundation" (*LH* 263).

To emphasise this common structure, Heidegger proceeds to unfold the putative objections to his work in six short paragraphs. Each of these begins with a statement of Heidegger's apparent view, followed by a statement of what 'people' take that view to imply.[16] Heidegger then ends each paragraph with a variation on the same question, in each case reiterating the conclusion drawn by the 'people': "For what is more 'logical' than that...?"(*LH* 263-4)[17] Each of these paragraphs takes up a different theme – 'humanism', 'logic', 'values', 'being-in-the-world', the 'death of God', and 'nihilism'. Heidegger first offers a general diagnosis of the misunderstandings pertaining to his

[16] It is perhaps significant that the term translated in the Cambridge University Press edition as 'people' is the neuter pronoun '*man*': '*befürchtet man*', '*meint man*', and '*entsetzt man sich*' are translated as 'people fear', 'people believe', and 'people are horrified' (*LH* 263), but could equally – indeed, perhaps more accurately – have been rendered as '*one* fears', '*one* believes', '*one* is horrified'. These translations would not only have mirrored the grammar of the German more closely, but would also have drawn attention to the possibility of a continuity between this stretch of the *Letter* and Heidegger's discussion of '*das Man*' in *Being and Time*. Earlier in the *Letter*, Heidegger had referred explicitly to this discussion, relating his claim that "language comes under the dictatorship of the public realm, which decides in advance what is intelligible and what must be rejected as unintelligible" to "what is said in *Being and Time* §§27 and 35 about the 'one'" (*LH* 242). The latter section concerns what Heidegger calls 'idle talk' – the 'average' kind of understanding that is unable to "decide what has been drawn from primordial sources with a struggle and how much is just gossip" (*BT* 212/H169). As will shortly become clear in the main text, Heidegger's diagnosis of the misinterpretations of his work will appeal to the idea that what 'speaking against' values, humanism, etc. might amount to – what kinds of opposition to these notions might be available and intelligible – has already been 'decided in advance' by the opinions of 'the people'.

[17] This phrase ("Denn was ist 'logischer' als dies, daß...?") appears six times in the course of two pages.

treatment of these themes, before providing a separate and self-contained clarification of his position on each of them in turn.

In each case, an opposition is assumed in advance, and falsely taken to be exhaustive of the available positions on the issue. But Heidegger's point is that to interrogate the terms in which an opposition is posed – or the interpretation given to those terms – might allow one to refuse to endorse either of the opposed sides. This is precisely the move Heidegger will make in the case of 'values'. Heidegger articulates the relevant opposition in the following way: "Because we are speaking against 'values' [*Werte*] people are horrified at a philosophy that ostensibly dares to despise humanity's best qualities. For what is more 'logical' than that a thinking that denies values must necessarily pronounce everything valueless [*wertlos*]?" (*LH* 263-4) Here, the choice is apparently between assigning value to 'humanity's best qualities', or consigning them to valuelessness. But this simply presupposes that the only way of making sense of the importance of something, of the way something might matter, or of the claim something makes on us, is in terms of the concept of a 'value'.

Heidegger thinks this presupposition is not just mistaken, but highly damaging. His dense and suggestive explication of his critique of this presupposition comes a few paragraphs later, and is worth quoting in its entirety:

> To think against "values" [*die Werte*] is not to maintain that everything interpreted as "a value" – "culture," "art," "science," "human dignity," "world," and "God" – is valueless [*wertlos*]. Rather, it is important finally to realize that precisely through the characterization of something as "a value" what is so valued is robbed of its dignity [*Würde*]. That is to say, by the assessment of something as a value what is valued is admitted only as an object [*Gegenstand*]

> for human estimation. But what a thing is in its Being [*das, was etwas in seinem Sein ist*] is not exhausted by its being an object [*in seiner Gegenständigkeit*], particularly when objectivity [*Gegenständlichkeit*] takes the form of value. Every valuing [*Werten*], even where it values positively, is a subjectivizing [*Subjektivierung*]. It does not let beings: be. Rather, valuing lets beings: be valid – solely as the objects of its doing [*das Objekt seines Tuns*]. The bizarre effort to prove the objectivity [*Objektivität*] of values does not know what it is doing [*weiß nicht, was sie tut*]. When one proclaims "God" the altogether "highest value," this is a degradation of God's essence. Here as elsewhere thinking in values is the greatest blasphemy imaginable against Being. To think against values therefore does not mean to beat the drum for the valuelessness and nullity of beings. It means rather to bring the clearing of the truth of Being before thinking, as against subjectivizing beings into mere objects [*zum bloßen Objekt*]. (*LH* 265, translation slightly modified)

The first point to make about this passage is that Heidegger explicitly responds to the two charges that 'people' might make against his critique. To reject talk of 'values' does not commit him to 'pronouncing everything valueless' – he is not 'beating the drum for valuelessness and nullity', not committing himself to a form of nihilism. Nor does he aim 'to despise humanity's best qualities' – or more literally: 'to abandon (or surrender) the highest goods of humanity to disregard' [*die höchsten Güter der Menschheit der Mißachtung preiszugeben*]. His claim seems to be rather that thinking in terms of 'values' forces us to construe those 'goods' in a particular – and particularly restrictive – way: only as objects for our own estimation. This prevents us from appreciating the proper 'dignity' of what we supposedly value.[18] Significantly, this 'thinking against' values is carried out in overtly thick normative vocabulary: valuation not only 'robs' what is valued of its dignity, but in the final analysis amounts to a 'degradation', even to 'blasphemy'. Thus, far from constituting a rejection of normativity as such, Heidegger's

[18] The German term, '*Würde*', is translated in the Cambridge edition as 'worth'. However, since 'worth' is closely associated in English with (monetary) value – precisely what Heidegger is here opposing – 'dignity' seems the preferable term. The choice of 'worth' also somewhat obscures the relationship to 'human dignity' [*Menschenwürde*], a notion explicitly mentioned at the start of the paragraph, and also discussed earlier in the *Letter* (see *LH* 251). Consequently, the implicit criticism of Kant's account of value and 'dignity' that I discuss shortly in the main text is also thereby obscured.

critique appears to turn on his sense that thinking in terms of 'values' is itself a kind of ethical failure, in a sense that we have yet to fully unpack.

But how can we make sense of a normative critique of 'values' in general? It would be natural to object here that our concepts of value and normativity bear an intrinsic relation to one another, and are not so easily pulled apart. If something is normative for us, it offers us a norm or standard and prescribes what we ought to do. But what are the sources of normativity, of this 'ought'? Certainly nothing in the world, no *fact* – for we know from Hume and Kant that no 'ought' could be derived from an 'is'. The source of the 'ought', the claim on us, must instead be the value possessed by something – for instance by the ends of our actions. As Heidegger himself puts the point in his 1935 lecture series *Introduction to Metaphysics*, "Something like an ought can emanate only from something that raises such a claim on its own, something that in itself has a *value*, and itself *is* a *value*. Values as such now become the ground of the ought" (Ga40: 221). It looks, then, as if our concept of value is so central to our understanding of normativity and the 'ought' that we cannot dispose of the one without dispensing with the other.

Nonetheless, this argument is not decisive. For, as Heidegger will be keen to stress, we have not yet clarified the notion of 'values' ontologically. Already in *Being and Time*, Heidegger had posed the question of what values amount to, and what we mean ontologically when we say that things are 'invested' with value (96/H68; 132/H99). Furthermore, Heidegger points out that the philosophical conception of 'value' itself has a history, suggesting in particular that it comes to prominence in modernity. It is

not a coincidence that the split between 'fact' and 'value', 'is' and 'ought', is typically traced back to Hume and Kant. It is part of a decisively modern shift, though in *Introduction to Metaphysics* Heidegger argues that it has its pre-history in Plato, and in a more general distinction between Being and the ought. In *Introduction to Metaphysics*, Heidegger thus provides what could be viewed as a brief genealogy of the concept of 'value', which sheds light on the critique presented in the *Letter*. Heidegger views Kant as an especially decisive figure in this genealogy, as he institutes a fundamental split between nature (the realm of what 'is', of beings) and morality – the realm of the 'ought', of 'values', and of validity. Although Heidegger does not mention Kant by name in the passage from the *Letter*, a discussion of some of Kant's views will shed further light on Heidegger's critique. There are two themes in particular that are worth drawing out. The first is the way that Heidegger repeatedly draws a connection between 'values' and the subject-object relationship. The second is the notion of 'dignity'. I shall consider each of these in turn.

Heidegger writes that "it is important finally to realize that precisely through the characterization of something as "a value" [*Wert*] what is so valued is robbed of its dignity [*Würde*]" (*LH* 265, translation modified). It is in this sense that on his view, talk of 'values' itself degrades – presumably in the sense of a lowering of rank rather than a lowering of price – precisely those beings to which we are accustomed to attributing the highest value: human beings, art, God. This thought informs his claim that "every valuing, even where it values positively, is a subjectivizing. It does not let beings: be [*sein*]. Rather, valuing lets beings: be valid [*gelten*] – solely as the objects of its doing" (*LH* 265). While the meaning of 'subjectivizing' here is not completely transparent, the

idea seems to be that talk of values relativizes the worth of that which is valued to its relation to the 'doing' of a subject. Rather than allowing beings to 'be' – and it would presumably be in this that their true dignity would consist for Heidegger – it transforms them into mere objects, each with a relative value for a subject. Another way to put this point might be to say that something valuable is always of value *for* something or other, of value *to* someone or other – either for some purpose, project, or 'doing', or for the subject whose purpose (&c.) it is. In a certain sense, then, talk of values is inappropriate to the beings we value, insofar as it stands to conceal their genuine dignity. It is the product of a view of the subject as *conferring* value on otherwise valueless things.[19]

For this reason, "the bizarre effort to prove the objectivity of values does not know what it is doing" (*LH* 265). This attempt at a proof is 'bizarre' because it is confused from the start, on two counts. It fails first of all to see that values are always related in some way to the positing of a subject, and are therefore not 'objective' in the intended sense (which is presumably something like 'subject-independent'); but secondly, it fails to see that the entire sphere of the 'objective' is (at least according to Heidegger) itself always already set up in opposition to a subject. Ontologically speaking, subject is always opposed to object; if we (humans) had not become subjects, had not posited ourselves as subjects, nothing like an object would ever show up for us. If such a claim appears odd, it should be borne in mind that for Heidegger, 'object' (*Gegenstand*) is to be absolutely distinguished from 'thing' (*Ding*) and being (*Seinde*). In the distinctively

[19] Consider in this connection Korsgaard's claim that "To treat anyone as an end in itself is to regard that person as one who *confers value* on the objects of his or her choice" (1996: 128, emphasis added). Recall also her notion of 'tethered' importance, discussed in Chapter One, Section II. For Korsgaard, value is presumably also something 'tethered' to valuing subjects. In a sense, I take Heidegger to agree with this latter point, but to disagree that 'dignity' has anything to do with 'value', so understood.

modern experience of beings as objects standing over against us (*Gegen-stand*), an experience of *things* in their 'thinghood' is occluded. This occlusion in turn prevents us from attending to the proper dignity, not *just* of human beings, but perhaps of things too; it is only in letting things be that they attain their proper dignity.[20]

It may be objected here that Heidegger's analysis, in claiming that all value is 'subjectivizing', fails to take account of our notion of an absolute or unconditional value. We might again call Kant to mind here, as an obvious example of a thinker who distinguished unequivocally between the absolute 'value' or worth of the human being, and the merely relative 'value' of the beings – including, as we saw in Chapter One, non-human animals – that humans may utilise as means for their ends. Sean Kelly has offered a response to that charge on Heidegger's behalf: although Kant opposes any relativisation of the value of the human, he nonetheless still conceives of the value of the human on the model of relative values – what Kelly calls an 'economic' model of values.[21]

If this is Heidegger's response to the objection at hand, it does not seem to be a convincing one. Consider again the passage from the *Groundwork* in which Kant distinguishes between price and dignity: "What has a price can be replaced with something else, as its *equivalent*; what on the other hand is raised above all price and therefore admits of no equivalent has [...] not merely a relative worth, that is, a price,

[20] Heidegger appeals to the example of a jug to make this point: "As the selfstanding [*Selbststand*] of something independent, the jug is distinguished from an object [*Gegenstand*]" that has been manufactured for a particular use (Ga79: 5).
[21] Sean Kelly, 'The Proper Dignity of Human Being: Later Heidegger and the Philosophical Tradition', paper delivered at the Post-Kantian European Philosophy Seminar, University of Oxford, Tuesday 19th November 2019.

but an inner worth, that is, *dignity*" (4:434-5). Although Kant does use the word 'value' [*Werth*, an archaic spelling of *Wert*] in both cases (the value of that which has a price and that which has dignity), the way he cashes out the difference between them makes it clear that he takes the two cases to be fundamentally different in kind. However we are to understand the notion of an 'inner worth' (or 'inner value') here, we are clearly not to understand it on the 'model' of a merely relative worth. Kant's suggestion seems to be that when we value dignity, we are valuing it in a way that is fundamentally irreducible to value understood quasi-economically as a 'price', as something replaceable with something else of equal value. This is a point that Christine Korsgaard has recently emphasised, in an objection to the common formulation that "people have *equal* value as ends in themselves." Korsgaard argues that on Kant's view, "the value we have as ends in ourselves is not so much 'equal' as 'incomparable.'"[22] When Kant suggests that the value of ends in themselves is unconditional, then, part of what he means is that we are not to treat that value on the model of an economic understanding of value – we are not to treat it, as Heidegger might put it, as calculable.

However, even if these considerations undermine Kelly's reading of Heidegger's argument, there is perhaps a deeper response available to Heidegger here. If Kant has given us a way to understand the 'inner value' of ends in themselves – and so of human beings – in non-economic terms, his explanation of inner value nonetheless makes it dependent on a subject. Kant is still guilty of a 'subjectivizing' of value, insofar as the 'dignity' of the human being is bound up with their status as a free rational being – as a *subject* with the capacity to set ends for themselves. On Korsgaard's reading of Kant's

[22] Korsgaard (2018: 65).

argument for the formula of humanity, Kant's point was that we cannot value anything else – we cannot perform the activity of valuation – without valuing ourselves. As we saw above, Heidegger's complaint was that characterising something as a 'value' deprives it of genuine dignity, since the valuing activity of a subject only "lets beings: be valid – solely as the objects of its doing." We are now in a position to see how this could even be true of the 'inner value' of ends in themselves on Kant's account. Their value is only guaranteed as a necessary precondition of their activity of valuation, so the 'dignity' of the human being is ultimately explained in terms of the (in this case reflexive) value-conferring activities of the human subject. Heidegger's criticism, then, is that Kant misconstrues the dignity of human beings, at the same time as obscuring altogether the idea that 'things' – and, most importantly, *Being* – might have a dignity of their own.

It seems, then, that Heidegger's thinking 'against' 'values' is not a rejection of the normative as such; indeed, it seems to be based primarily on normative considerations, insofar as talk of 'values' is opposed in part because it 'degrades' beings and does not allow them their proper 'dignity'. Thus, there is no contradiction between the rejection of 'values' and the unmistakable normative urgency of Heidegger's later writings. On the contrary, Heidegger's critique of 'values' is itself implicitly normative. There is a strong implication that it would be *better* for us to 'let beings be' than to merely 'let beings be valid'. Indeed, he appears to be suggesting that it is precisely in letting beings be that *our* 'proper dignity' lies. It is worth pointing out, however, that on such a conception humans would not be dignified *as such*. Human dignity would consist in a possibility, one that may or may not be realised. Heidegger's account would thus

contrast with more Kantian accounts of dignity, according to which human dignity marks us out as ends in ourselves. The latter notion appears to be foreign to Heidegger's conception of the proper dignity of human beings.

In this section, I have argued that Heidegger's critique of 'values' is consistent with the apparent normative urgency of his later work. Part of the impression of normative urgency comes from Heidegger's near-constant use of – at least apparently – 'thick' ethical concepts, rather than abstracted ones such as 'value'. In the final part of Section II, I turn to some particular examples of such thick vocabulary in Heidegger's *Letter*, in the hope that elucidating these terms will shed further light on Heidegger's understanding of the inhuman, and its relation to the ordinary moral understanding of the same concept.

C. The Question of Evil

Late in the *Letter*, Heidegger makes some brief and cryptic remarks about 'healing' and 'evil':

> And yet thinking never creates the house of Being. Thinking conducts historical eksistence, that is, the *humanitas* of *homo humanus*, into the realm of the upsurgence of healing [*Heilen*].
> With healing, evil appears all the more in the clearing of Being. The essence of evil does not consist in the mere baseness [*Schlechtigkeit*] of human action, but rather in the malice of rage. Both of these, however, healing and the raging, can essentially occur in Being only insofar as Being itself is in strife. (*LH* 272)

The passage begins with some ideas that should by now be familiar: the house of Being is *not* a human creation; rather, thinking (in the more primordial sense) is a kind of conduit through which we first attain our essence, '*humanitas*', through a renewal of our

relationship to Being. However, from there, the questions multiply: What is the 'upsurgence of healing'? If – as seems quite natural – we are to hear salvific connotations in this phrase, how can it be that 'with healing, evil appears all the more'? What is the difference between 'evil' and 'the essence of evil'? If the latter does not consist in the 'mere'[!] baseness of human action – and *Schlechtigkeit* might just as well be rendered 'wickedness' or 'badness' here – what *does* it consist in? 'The malice of rage' – but how are we to hear 'malice' and 'rage' here in a way that does not refer back to (base, wicked, 'bad') human action? Finally, how is the 'essential occurrence' of healing and raging enabled by Being, 'insofar as it is in strife'?

Passing over the first question for the time being, let us consider the second question. There is an ambiguity here regarding the word 'appears'. On one reading, it sounds as if Heidegger is saying that healing somehow causes (or brings about) the appearance of evil. If we understand Heidegger's claim in this way, and moreover preserve for the time being the 'positive' connotations of 'healing' and the 'negative' connotations of 'evil' (and surely few words have more negative connotations than this one!), the claim looks puzzling, even paradoxical. However, there may be another interpretation of 'appears' available here. Perhaps the idea is rather that as our 'historical eksistence' is conducted into 'the realm of the upsurgence of healing', all the more will 'evil' appear in the sense of 'becoming visible'. The idea would be that it is only when we enter into this healing that we can first *become aware* of an evil that was already in some sense 'raging'. This provides a somewhat less counterintuitive construal of Heidegger's statement.

What of the third question? It is important to note here that for Heidegger the word 'essence' has a rather different sense than it does in traditional metaphysics. We can first clarify it negatively by noting that what is *not* meant is '*essentia*': essence in the sense of 'what something is', or that which is common to all entities of a particular kind, in virtue of which they are the entities they are. For later Heidegger, however, 'essence' seems to be something dynamic or processal, something that 'unfolds' temporally. To convey this, Heidegger derives a verb, *wesen*, from the German noun *das Wesen* (essence). This verb is often translated (as in the passage quoted above) as 'to essentially occur', and sometimes as 'to essence'.[23] So the essence of evil is something that itself occurs, that is 'in process'. Often, for instance in his discussion of (the essence of) technology, Heidegger will argue that the 'essence' of something is also in some way prior to it, so that 'technology' in the everyday, modern sense is something like a local manifestation or consequence of the essence of technology understood as an event or epoch in the history of Being. Returning to the question of the relationship between evil and the essence of evil, then, we can assume that Heidegger is here establishing something like a priority relation: the suggestion seems to be that it is *not* primarily through the 'baseness' of human actions that evil comes about, but, on the contrary, that 'evil' in the Heideggerian sense – the malice of raging – is an event in the history of Being. It remains to be seen, then, how 'human' evil is related to this Being-historical evil.[24]

[23] The latter translation has the advantage of better matching the German syntax, though both translations sound awkward in English – as their original no doubt does in German.
[24] I use 'Being-historical' (*Seinsgeshichtliche*) here, as Heidegger does, as an adjectival form of 'history of Being' (*Seinsgeschichte*). The 'Being-historical' sense of a term is, roughly, its sense insofar as it designates or is related to an event in the history of Being.

Unfortunately, there are few other mentions of these notions in the *Letter* itself. On the following page, we are told that "To healing Being first grants ascent into grace; to raging its compulsion to malignancy" (*LH* 273). However, this is not so much a clarification as it is a reiteration of the claim regarding the provenance of the 'essential occurrence' (namely, Being), and a confirmation that in discussing 'the raging' Heidegger is no longer talking about a human doing, but rather something 'granted' by Being itself. Finally, the reference to 'malignancy' here recalls an earlier passage, in which Heidegger surmises that "Perhaps what is distinctive about this world-epoch consists in the closure of the dimension of the hale [*des Heilen*]. Perhaps that is the sole malignancy [*Unheil*]" (*LH* 267). This sentence occurs at the close of a discussion of the essence of the holy [*das Wesen des Heiligen*], and foregrounds an etymological connection between the holy [*das Heilige*], the healing (here rendered 'the hale') [*das Heilen*], and the 'malignancy' or harm [*das Unheil*]. If this 'sole' malignancy is later connected to 'the raging' (and therefore the essence of evil), we can perhaps conclude further that there is a connection between the raging and the 'closure' of the dimension of the healing. But, in line with our conjecture above, we can only truly recognise this evil, the raging, for what it is when we have been granted access to 'the upsurgence of healing' – and this is somehow brought about through thinking.

Having examined in a cursory manner the explicit statements about evil in the *Letter*, we can draw the following tentative – and surprising – conclusion: Heidegger connects 'the essence of evil' not so much with the inhuman(e), the evil actions of which humans are capable, but rather with the radically *non*-human – with Being. In the *Letter*, the essence of evil is said to 'rest in the malice of rage', which can only 'essentially occur in

Being.' What is still unclear is what connection, if any, Heidegger perceives between evil in this Being-historical sense and evil in the 'ordinary' moral sense: the sense connected with human (all too human) 'inhumanity'. To answer this question, I turn now to a more extensive discussion of evil in the third of Heidegger's *Country Path Conversations*, written at the close of the Second World War and titled 'Evening Conversation: In a Prisoner of War Camp in Russia, between a Younger and an Older Man'.

III. Evil and the Abandonment of Being: A Troubling Conversation

A. The argument of the Evening Conversation

Given that the *Letter* (December 1946) was composed soon after the third conversation (May 1945), and given the close affinity in the phrasing of Heidegger's remarks on evil in the two texts, the latter seems likely to offer clarification regarding the obscurities of the former.[25] Some common themes emerge in the conversation's opening pages. To begin with, both texts connect evil with healing. In Section II, I argued that Heidegger's puzzling claim in the *Letter* that "with healing, evil appears all the more in the clearing of Being" (*LH* 272) is best read as suggesting not that 'healing' brings evil about, but rather that we gain insight into evil (and its essence) when we are 'conducted into the realm of healing'. The movement of the *Evening Conversation* lends support to this

[25] Drawing on a dialogue rather than an essay or lecture naturally raises methodological questions. Most pressingly, one may wonder which of the views expressed in it, if any, correspond to Heidegger's own. However, while this issue obviously warrants some caution, in the case of the *Conversations* it is usually fairly obvious which of the views expressed by the (somewhat archetypal) 'characters' are endorsed by Heidegger himself. The conversations are couched in consummately Heideggerian vocabulary, mostly drawn from the *Contributions* and adjacent writings, and can be understood as part of Heidegger's continuing attempts to "bring [the] manifestation [of Being] to language and to preserve it in language through [his] saying" (*LH* 239).

reading, by illustrating precisely that process. The impetus of the conversation is the Younger Man's experience of being "suddenly overcome by something healing" (Ga77: 132). The Older Man remarks that to understand the nature of this healing, we "would have to know what is wounded in you" (Ga77: 133), and the attempt to express 'what is wounded' leads directly into a discussion of the surrounding destruction of the war, and the 'devastation' underlying it. The discussion of the essence of evil is set in motion by the Older Man's claim that this "devastation of the earth and the annihilation of the human essence that goes with it are somehow evil itself" (Ga77: 133). What is suggested by the shape of the dialogue, then, is that an experience of genuine healing can lead to clarity on the nature of 'the wound' and of evil itself, both of which (as we shall see) are for Heidegger intimately connected with an event in the history of Being – namely, the abandonment of Being [*Seinsverlassenheit*].

A further resonance between the two discussions finds expression in Heidegger's claim from the *Letter* that "The essence of evil does not consist in the mere baseness [*Schlechtigkeit*] of human action, but rather rests [*beruht*] in the malice of rage" (*LH* 272, translation modified). Heidegger seeks here to distinguish between *moral* evil ('baseness') and the essence of evil, or evil understood in an essential or Being-historical sense. He elaborates on this idea at greater length in the *Evening Conversation*, a recurring theme of which is the relation between the essence of evil and two sets of moral 'phenomena': instances of moral evil, and possibilities of moral evaluation. Regarding the first set (we shall return to the second later), Heidegger introduces his discussion of the 'devastation' that he identifies with evil as follows:

> The devastation [...] has not, after all, existed just since yesterday. And it is not exhausted by what is visible and tangible. It can also never be accounted for by an enumeration of instances of destruction and the obliteration of human lives, as if it were only the result of these. (Ga77: 133)

The 'only' in the final sentence need not be read as diminishing the significance of these 'tangible' events. Rather, Heidegger's idea is that the devastation somehow exceeds them. Indeed, he argues that 'devastation' in his sense actually *underlies* them: "devastation is in no way a consequence of the World War, but rather the World War is for its part *only a consequence of the devastation* that has been eating away at the earth for centuries" (Ga77: 136, emphasis added; see also 142, and the postscript *in propria persona* at 157). But, anticipating the phrasing of the *Letter*, the Younger Man tells us that the evil of this underlying devastation is to be understood not as "what is morally bad, and not what is reprehensible, but rather [as] malice" (Ga77: 133). As the Older Man points out, this contrast is initially highly puzzling: how are we to understand malice in a way that is not correlated with the morally bad? The statement that 'evil is malice' is especially odd in German, where "as the name says, malice [*das Bösartige*] is of the nature [*Art*] of evil [*Bösen*], and is an outflow of it" (Ga77: 134).

Heidegger's answer is that we can give a "profound sense" to this statement, if we "think of malice on the basis of something other than morality" (Ga77: 134). This 'something other' is connected, as it is in the *Letter*, with 'rage':

> Malice is insurgency, which rests in furiousness, indeed such that this furiousness in a certain sense conceals its rage, but at the same time always threatens with it. The essence of evil is the rage of insurgency, which never entirely breaks out, and which, when it does break out, still disguises itself, and in its hidden threatening is often as if it were not. (Ga77: 134)

This elaboration on 'malice' may seem to replace one obscure term with several more, getting us no further. Nonetheless, connections can be perceived between this passage and those quoted above concerning devastation and destruction. In particular, the idea that devastation exceeds its tangible manifestations is alluded to again here: rage 'never entirely breaks out', that is, its expressions never exhaust it; it 'disguises itself', 'is often as if it were not', thus remaining hidden 'for centuries', unnoticed beyond its 'visible and tangible' signs. These ideas raise two further questions: what does this 'rage' consist in, and how does it disguise itself?

While the text is hardly transparent on these points, the Older Man speculates that the 'rage' of malice might be "something pertaining to the will," leading the Younger Man to make the "audacious" suggestion that "the will itself is what is evil" (Ga77: 134). Part of the audacity of this claim is that we usually see evil as a *possibility* of the will – one that is contrasted with the good will and morality. To identify the will *itself* with evil seems to preclude the possibility of a 'good will' that is not somehow tainted by evil. Heidegger indeed hints at this possibility: "it could be the case that even morality [is] only a monstrous offspring [*Ausgeburt*] of evil" (Ga77: 135). This claim leads into a discussion that both acknowledges and disavows Nietzsche as a possible precedent. According to Heidegger, Nietzsche posited affirmation of the will to power as an alternative 'beyond good and evil'. Such an affirmation is a futile attempt to escape this evaluative opposition, however: "Assuming [...] the will itself is what is evil, then the realm of pure will to power is least of all a 'beyond good and evil'" (Ga77: 135). On the contrary, the affirmation of will to power (which in Heidegger's view becomes

dominant in modernity) only serves to entrench evil in the Heideggerian sense, feeding the devastation.

This discussion brings out once again Heidegger's distinction between *moral* evil and the essence of evil. A parallel can be found in his essay on technology, where our usual anthropocentric understanding of modern technology that views it in essentially instrumental terms is supplanted by an understanding of the essence of technology as an objectifying mode of revealing, of disclosing entities. Further, in a claim that resonates with ideas in the *Evening Conversation*, Heidegger stresses that "The threat to man does not come in the first instance from the potentially lethal machines and apparatus of technology" (*QCT* 28). Rather, the 'danger' that technology represents is the annihilation of the human *essence* (cf. Ga77: 12-16, 21, 57, 103) – the possibility that we will be unable to "experience the call" of the truth of Being. This is not just a structural similarity between the argument of the two texts; rather, technology and the devastation belong to the *same* epoch in the history of Being, rooted in what Heidegger terms the abandonment of Being: "The Being of an age of devastation would then consist precisely in the abandonment of Being [*Seinsverlassenheit*]" (Ga77: 138).

The sense Heidegger attaches to this phrase can be illuminated by a closer look at the meaning of 'devastation'. For Heidegger, "'Devastation' [*Verwüstung*] means [...] that everything—the world, the human, and the earth—will be transformed into a desert [*Wüste*]" (Ga77: 136). This is a desert in a Being-historical sense, rather than a 'geographical' one; 'world' and 'earth' have what the early Heidegger would have referred to as an 'ontological' signification, so their 'desertification' is not merely a

product of 'ontic' destruction. On the contrary, this desert "allows for nothing that emerges of itself, in its emergence unfolds itself, and in unfolding calls others into a co-emerging" (Ga77: 137). As the translator points out (Ga77: 137, fn. 2), these terms allude to Heidegger's notion of *phusis*, the original Greek experience of Being. The suggestion, then, is that in the 'age of devastation' we can no longer have an experience of Being; we "remain abandoned by Being" (Ga77: 137). This does not, however, mean that in this epoch of abandonment beings 'are not,' or lose their Being: the abandonment "nonetheless still lets beings be" (Ga77: 138). What is lost is an experience of Being *as such*, what Heidegger sometimes terms 'Beyng' [*Seyn*], using an archaic spelling to mark the difference between 'Beyng' and the 'Being of beings'. We can no longer experience what the *Contributions* calls Beyng's "giving self-withdrawal" (Ga65: 196, §131). This is why the essence of technology is intimately connected with this 'abandonment': the totalising domination of the technological mode of revealing threatens to occlude both the possibility of other modes of 'truth', of other 'gifts' from Beyng, *and* the very fact that technology in the essential sense itself 'comes from' Beyng.[26]

This feature of the abandonment of Being returns us to our original question regarding the relation between moral evil and the essence of evil. For Heidegger, human beings "instigate and sustain [...] consequential phenomena of the devastation, though *never* the devastation itself" (Ga77: 136, emphasis added). Devastation does not originate in human acts; on the contrary, "malice, as which the devastation occurs, may very well remain a basic trait of Being itself [because] the devastation rests in the abandonment

[26] For an account of how this occlusion takes place, and why for Heidegger it represents the 'supreme danger', see Thomson (2011: 196-200, especially 199).

of beings by Being, [which] comes forth from Being itself" (Ga77: 139).[27] In keeping with his rejection of anthropocentrism, Heidegger stresses that devastation, while it essentially concerns human beings, is not created by them, but rather has its provenance in Being; it is an occurrence in the history of Being. But for that reason, we are barred from "regarding this thought, that evil would dwell in the essence of Being, as 'pessimistic' or in any way evaluating it" (Ga77: 139). Indeed, Heidegger repeatedly stresses the inappropriateness of moral "indignation", "superiority", "haughtiness", and "false passion", and the mistake of "morally evaluating the devastation" (Ga77: 133-136). Nor are variations on this claim limited to the *Evening Conversation*. Heidegger makes a similar point in his interview with *Der Spiegel*, 'Only a God Can Save Us', when he suggests that it is not his place to "preach and to impose moral judgment." Indeed, he claims that this would be "contrary to the meaning of the task of thought" (OAG 110).

Heidegger offers several reasons for this suspicion against a 'moral' response to the devastation, to 'evil' in the essential sense. The first is that it makes, if you like, a category mistake – it seeks to evaluate in moral terms something that is not the work of moral agents. The devastation is "an event that prevails outside of human guilt and atonement" (Ga77: 140). Beyng is not a being, and is certainly not a morally culpable human being. The second is that a moral response will be utterly ineffective: "moral superiority is not in a position to grasp, much less abolish or even mitigate, evil" (Ga77: 134). There is scope for a human response to evil, but it consists in preparing ourselves

[27] For an account of Heidegger's claim that evil might be conceived as a 'basic trait of Being itself' that explains its relation to ideas in the German idealist tradition (especially the work of Schelling), see Davis (2007: 289-98).

for a thoughtful relation to Being, in what Heidegger alternately characterises as thinking, 'releasement', and, in the *Evening Conversation*, 'waiting'; it does *not* consist in "the setting up of a morally grounded world order" (Ga77: 139). On the contrary: Heidegger's third – and perhaps most controversial – reason for resisting a 'moral' response is that moral programmes only serve to entrench the devastation further:

> Older Man: [That] which inflicts the suffocating hides itself behind something insidious, something which announces itself in the form of the purportedly highest ideals of humanity: progress, unrestrained escalation of achievement in all areas of creating, equal employment opportunities for everyone, and above all the allegedly highest rationale—the uniform welfare of all workers.
> Younger Man: What is really devastating, and that means what is malicious, consists here in the fact that these goals for humanity lead the various realms of humanity to become obsessed with devoting everything to their realization, and so with unconditionally driving the devastation onward while increasingly reinforcing it in its own consequences. (Ga77: 136)

I shall raise some critical questions concerning this passage later. For now, I simply wish to stress its relation to the aforementioned claim that evil 'disguises itself, and in its hidden threatening is often as if it were not.' We can now see more clearly how this self-disguising takes place: precisely through the positing of 'goals for humanity' that are opposed to destruction and therefore *apparently* to devastation. These goals, in Heidegger's view, are just as much an expression of the 'raging' of the will as the moral evils of the war are. A passage from the *Contributions* puts this in stark terms. Heidegger writes that for Nietzsche, "*Nihilism* means [...] that we lack all *goals*." However, "the genuine nihilism" is rather

> the unwillingness to acknowledge the lack of goals. [...] Therefore precisely where one believes one again has goals [...] resides the greatest nihilism, the deliberate turning of a blind eye to human goal-lessness, the [...] avoidance of any goal-setting decision, the dread of all decisive domains and of their opening. The dread of Beyng was never as great as it is today. (Ga65: 109, §72)

A reference to *Volk* elsewhere in the full passage suggests that Heidegger's critique here is partly aimed at National Socialism, with which he grew increasingly disenchanted in the late 30s. But the critique clearly has implications that transcend the differences between political programmes. For Heidegger, our very confidence in our goals, including 'progressive' ones, closes us off from an awareness of our genuine 'goal-lessness' – which, as the final sentence suggests, is intimately connected with the abandonment of Being. The end result of remaining within the moral sphere, driving towards the achievement of a 'moral world order' as our 'highest ideal', is that "the abandonment by Being is hidden away", made "unassailable" (Ga65: 95, §57).

B. Two Questions for Heidegger

With this elaboration of Heidegger's discussion of evil in hand, it is now possible to articulate two interrelated critical questions regarding his strategy and overall normative project. The first is a theoretical question concerning the account's 'explanatory power', the second a normative question concerning its implications. I shall begin with the theoretical question: in what sense is it the case that "the destruction could not even begin" unless "the devastation was already at work" (Ga77: 142)? Put more generally, what precedence do events in the history of Being have over particular, 'ontic' historical phenomena, such that these phenomena can be 'consequences' of them (cf. Ga77: 136)?

Presumably, Heidegger's answer will have to be given at a very general level, and in the case of moral evil it will be given in terms of the will; the ascension in modernity of what Heidegger elsewhere calls the "will to will" is somehow meant to underlie the

catastrophic events of the 20th century (Ga5: 176).[28] But can such an abstract explanation suffice to account for historiological events in their specificity? This question is especially pressing given Heidegger's view that phenomena that are apparently diametrically opposed at the ontic level may derive from the same Being-historical source. Not only morally *evil* acts, but *morality itself*, including attempts to institute a moral 'world-order', programmes of welfare, and human rights are alike suggested to be 'monstrous offspring' of evil (Ga77: 135-9). But since we surely cannot afford to remain neutral regarding which of these 'opposites' is to be preferred, Heidegger ought to make some suggestion as to *why* a Being-historical process that apparently makes *both* possible sometimes manifests as one and sometimes as the other.

It might be replied that this objection is misaimed: Heidegger owes us no explanations of this kind, for he is doing *philosophy*, rather than ordinary history ('historiology') or any ontic human or social science. Heidegger himself maintains that he "in no way means to furnish an incidental contribution to sociology" (*LH* 242). While that comment concerns in particular the discussion of '*das Man*' in *Being and Time*, it also applies quite generally. However, despite such warnings, it remains the case that as soon as Heidegger intimates a relationship between particular 'historiological' phenomena and his history of Being, it is difficult to see how he can avoid offering explanations that in *some* way compete with ordinary historiological ones, and he therefore owes us an explanation of the relationships between the two 'levels'.

[28] For an account of the development of Heidegger's views on the will and his critique of it, see Davis (2007), especially Chapter 7.

This point can be clarified with an example from the *Letter*. There, Heidegger claims that "What Marx recognized in an essential and significant sense [...] as the alienation [*Entfremdung*] of the human being has its roots in the homelessness of modern human beings" (*LH* 258, translation modified). He suggests that what constitutes the 'superiority' of the Marxist account is that "Marx by experiencing alienation attains an essential dimension of history" (*LH* 259, translation modified), namely an intimation of the 'homelessness' of humanity. This latter must be understood "in terms of the history of Being" (*LH* 258). The homelessness of the human being in modernity consists (at least partly) in our alienation from language, properly understood as "the house of Being[,] in [whose] home human beings dwell" and consequently in our distance from our essence (*LH* 239); it thus ultimately relates back to the abandonment of Being (though this is not explicit in the *Letter*). The essential point here is that our 'homelessness' is intimately connected with the modern epoch in the history of Being.

Since Marx himself obviously did not understand *Entfremdung* in this manner, Heidegger's comments on Marx are not to be read as an endorsement of Marxism as such – indeed, elsewhere Heidegger's antipathy towards Marxism is explicit and vehement (see e.g. Ga40: 41, 51-2, 54). Rather, Heidegger is making a claim parallel to what he says concerning Nietzsche and nihilism. If Nietzsche "caught sight of the appearances of nihilism [but] did not yet grasp its essence" (Ga77: 138), then Marx caught sight of the appearances of homelessness (viz. alienation), but did not himself understand it in terms of the history of Being.[29] However, if this is the claim, then

[29] Elsewhere, Heidegger claimed that "The essence of nihilism, which finds its ultimate consummation in the domination of the will to will, resides in the oblivion of Being" (Ga9: 319). Since for Heidegger the 'will to will' is the historical successor of the 'will to power' that

Heidegger's understanding of alienation (a modern historiological phenomenon) as *rooted* in homelessness (a Being-historical one) is in some sense meant to be an alternative to Marx's, and indeed a deeper one. The Marxist account, by remaining at the level of a materialist theory of economic history, is from Heidegger's perspective only superficial. However, such a critique can only be compelling if there is a demonstrable relation between the 'deep' history of Being and the historiological phenomena that other accounts understand only 'superficially'.[30]

Heidegger's account does not seem to have the resources to explain the specificity of the phenomena it claims are rooted in Being-historical developments. For instance, an account of alienation that tears it from its relation to capitalism's development, modern working conditions, and massive economic inequality is likely to efface important features of that phenomenon. An account that contains *no* discussion of these historical developments will pass over aspects of alienation that are highly normatively significant. Heidegger's history of Being is extraordinarily ambitious, attempting to show that all of Western history stands in the shadow of the metaphysical mode of questioning initiated by the Pre-Socratics and exhausting itself with Nietzsche, its last

---

Nietzsche championed – the consummate form taken by the will in the age of technology – the implication here is not just that Nietzsche failed to grasp the essence of nihilism, but that his philosophy was expressive of it, or even inadvertently fostered it or brought it to fruition.

[30] Richard Wolin writes that "the dimension of ontic life or everyday concretion would seem to fall beneath the threshold of Heidegger's ontological vision. And consequently, his [ontological] category of 'historicity' would never be capable of accounting for the [ontic] events of 'real history'"; see Wolin (1993: 157). But so put, this criticism is not decisive. It relies on an overly simplistic understanding of the relationship – and the distinction – between the ontological and the ontic. It is not that ontic phenomena simply fall 'beneath the threshold' of Heidegger's interest; in fact, he constantly alludes to ontic phenomena as manifestations or particular modalities of the epochs he discerns in the history of Being. There is no general problem with wanting to uncover the ontological enabling conditions of ontic phenomena. The problem instead arises at the level of specific phenomena, whose exact relationship to the history of Being Heidegger often leaves mysterious.

consequence modern technology and nihilism. But its very scope makes it hard to see how Heidegger's history can be anything but merely suggestive and vague at the level of specific 'ontic' phenomena, rather than genuinely explanatory.

Returning to our original question, the same point can be raised in relation to Heidegger's suggestions concerning the war. It is surely important that some consequences of 'the devastation' (the abandonment of Being) include "destruction and the obliteration of human lives" (Ga77: 133), even if these are not synonymous with or 'exhaustive' of that event in the history of Being. But Heidegger seemingly has no interest in specifying *how* these events are made possible by the devastation, or in making any suggestions as to how what we might see as its more positive expressions are to be fostered. On the contrary, as we saw above, he suggests instead that "what is really devastating" is our obsession with apparently positive goals of 'progress' and welfare, which "unconditionally drive the devastation onward while increasingly reinforcing it in its own consequences" (Ga77: 136).

It is left ambiguous *which* consequences are 'increasingly reinforced' by this entrenchment. If the suggestion is that ontically destructive consequences are reinforced by our pursuit of progress, Heidegger owes us an explanation for why that should be. A partial answer is that this obsession covers up the danger to our essence that modernity represents. It mistakes the elimination of destruction and injustice for the elimination of devastation itself; thus, the "process of devastation will [...] not be warded off, much less ended with the setting up of a morally grounded world order" (Ga77: 139). Perhaps the idea is that if the 'process of devastation' continues, the

destruction will too. But more would need to be said if that suggestion is to be compelling. Even if we grant the strong claim that devastation is a necessary condition for destruction (and it is not clear that we *should* grant it), it would not follow that it is a sufficient condition for it. Some actions (such as safeguarding human rights) that for Heidegger remain within the devastation's scope might nonetheless make destructive moral evil much less likely.

A more radical possibility is that Heidegger views *all* the consequences of the devastation as problematic, because they all (whether 'good' or 'evil') equally express the 'annihilation of the human essence'. This connects to a second critical question that Heidegger's discussion invites: what are its normative implications, and where does Heidegger stand regarding our familiar (moral) evaluative resources? It does not *follow* from Heidegger's dismissal of the efficacy and appropriateness of moral evaluation of 'devastation' (*qua* abandonment of Being) that we have no right to moral evaluation more generally. Nonetheless, a troubling aspect of his account is that it apparently leaves this question open. If for Heidegger the "supreme danger" is technology (*QCT* 26-8), and the attendant possibility that our essence will be 'annihilated', and if the genuine 'wound' is "the withdrawal of the human existence [*Dasein*] that rests on the ground of thinking" (Ga77: 142), how does the normative demand that we prepare ourselves for genuine thinking (thus attaining our essence) relate to the other claims placed on us? Are these latter simply subordinate to the former, somehow less urgent? Does the claim that our familiar distinction between moral good and evil springs *from* evil in the Being-historical sense throw this distinction into question, so that we cannot rely on it until our proper relation to Being is restored?

Much of the initial sense of normative urgency in Heidegger's later texts comes from the connotations of the terms he uses to describe our situation: 'devastation', nihilism, the 'supreme danger' of technology, and the language of woundedness and healing. But if it turns out that *none* of these terms have the significations we expect – if the 'annihilation' of the human essence is repeatedly distanced from the 'obliteration' of actual human lives, if the 'devastation' reigns as much where rights and welfare are promoted as it does where war and destruction are rife – it becomes less than clear why the 'devastation' is something we should *want* to overcome.

At the very least, there is something disquieting in remarks such as Heidegger's claim that "this waiting people" – soon identified as 'the Germans' – is "endangered like no other" (Ga77: 152). Heidegger is quick to distance himself from a 'nationalistic' interpretation of his view that "by becoming those who wait, we first become German" (Ga77: 153) in what the translator Bret Davis describes as a critique "obviously aimed at the racist nationalism of Hitler's National Socialism" (Ga77: 153, fn. 13).[31] But this critique does not go far enough, and is not developed in the right direction. There is no clear justification for retaining the goal or vocabulary of 'becoming *German*' if that latter term has no nationalistic basis. Speaking 'Being-historically' about 'the German people' is at best misleading. Furthermore, the phrase 'endangered like no other', applied to any 'people' specified 'Being-historically' seems astonishingly inappropriate in an historical context where the Nazi party were committing systematic genocide. One may ask with some justification what right Heidegger has to the term 'danger' here, if he is intent on repeatedly distancing it from the atrocities taking place around him.

[31] Cf. the critical remarks concerning nationalism in the *Letter* (*LH* 260).

Conclusion

In order to bring out this problem more sharply, let us return to the question of Heidegger's view of the 'inhuman'. If to be inhuman is to be 'outside the essence of the human' (cf. *LH* 244), then our understanding of the inhuman is inextricable from our conception of the human essence. As we saw, Heidegger maintains that his thinking "in no way advocates the inhuman" (*LH* 263), and indeed, it should be clear that in *one* sense of that term, he stands absolutely opposed to it: the central and guiding aim of his later philosophy is to return us to our essence and to our relationship with Being through the cultivation of 'thinking'. However, the relation of his normative claims to the ordinary moral sense of the 'inhuman' – call it the humanistic sense – is ultimately ambiguous and troubled. Clearly Heidegger does not *endorse* the inhuman in this sense; the tone of the *Evening Conversation* suggests an embittered, mournful awareness of the destruction of the war. Nevertheless, by moving beyond moral good and evil, and questioning the efficacy of moral responses to the devastation of modernity, his account does seem to efface certain boundaries between the 'inhuman' and its opposite – for instance, between the inhumanity of the actions of the Nazi regime and the attempts elsewhere to secure and promote human rights.

It cannot be insignificant that *both* the 'inhuman' and the 'human' in the humanistic sense are 'inhuman' in Heidegger's 'more essential' sense – that for Heidegger "the devastation reigns also and indeed precisely there, where country and people have not been affected by the destruction of the war [...] where human rights are respected, where civil order is maintained" (Ga77: 139). The text of the *Evening Conversation* evinces rather *more* suspicion of the advance of humanistic goals than it does of overt

acts of moral evil, insofar as for Heidegger the devastation is more insidious when it is covered over by apparent 'progress'. But can we really regard this 'progress' as merely apparent, or as awaiting validation by our renewed relation to Being? At the very least, we must maintain an unequivocal recognition of the difference between Auschwitz and the ideals that Heidegger finds so suspicious. This is what Heidegger does not ever explicitly provide, and it is extremely troubling that in his opposition to the inhuman in a 'higher' sense, he directs his critique against resources that are a crucial safeguard against inhumanity, evil, in the 'merely' moral sense.

Chapter Three
'A New Basis for Humanism'?
Judith Butler on Vulnerability, Dependency, and the Body

There is an apparent tension in Judith Butler's work between, on the one hand, a critique of forms of dehumanisation, and, on the other, a suspicion regarding both 'humanism' as a political or philosophical position, and conceptions of 'the human' as a normative ideal. If 'dehumanisation' is something we ought to denounce, to expose, and to subject to critique (as it surely is), then presumably the notion of 'the human' must be playing *some* normative role. For the charge of 'dehumanisation' to make sense, we must presuppose both that there are ways we ought to treat other human beings and acknowledge them *as* human, and that it is possible to fail so to treat them.

Nonetheless, Butler thinks we cannot uncritically take 'the human' to be a normatively innocent term. They argue that substantive conceptions of 'the human' have themselves historically been – indeed still are – vehicles for the dehumanisation of certain populations, and are implicitly operative in the justification and normalisation of violence towards those populations. One way of resolving the tension, then, would be to reject normative conceptions of the human that operate in this manner. Butler articulates a 'new bodily ontology' that might be thought to do exactly this: in opposition to the picture of the human as a free sovereign individual, Butler stresses our vulnerability, our precariousness, and our interdependency. Indeed, Butler indicates the possibility of just such an interpretation of their work, although with severe reservations: "By insisting on a 'common' corporeal vulnerability, I may seem to

be positing a new basis for humanism. That might be true, but I am prone to consider this differently" (*PL* 42-3).

Butler's second sentence expresses a certain ambivalence: they register the possibility of developing their claims in this direction ('that might be true'), but mark their resistance to doing so. But just how strong is that resistance? In expressing their inclination in this way, Butler perhaps intends to leave room for the opposite inclination, and for an interpretation that connects their account of vulnerability with overtly humanistic aims. This is the interpretation I shall seek to defend in this chapter, by arguing that a more critical humanism can take its point of departure from Butler's ontology of the body. In thus resolving Butler's ambivalence, and the tension identified above, I may be advancing a reconstruction of Butler's position, rather than providing straightforward exegesis. Nonetheless, I hope to show that the resources for such a project can be found in Butler's own work.

I. Humanity, Universality, and Recognition

Let us consider first Butler's worries concerning the uses and effects of "certain exclusionary conceptions of who is normatively human" (*PL* xiv-xv). To be 'normatively human' is to *count* as human in a double sense: to be 'counted' as a human being, and thus to 'count' normatively, to be seen to *matter*. If conceptions of the human are 'exclusionary', they fail to 'count' some human beings as human: they fail to recognise their humanity, consigning them to the sphere of the less-than-human, the inhuman, or even the non-human. This failure typically leads to a (more or less extreme) suspension of the rights that ought to be accorded to them, and becomes the occasion for violence

of various sorts. If one is not recognised as human, one is unlikely to be treated humanely.

One of Butler's worries in this regard is about philosophy: philosophers' substantive attempts to produce an account of what human beings are, to provide a philosophical anthropology, or an account of our species-being, have tended to be parochial and exclusionary. They have failed to do justice to human difference, instead setting up 'false universals' – treating as a universal normative standard an account of the human that is highly historically and culturally specific, and violently imposing local standards of what it is to live a flourishing human life (in the context of imperialist colonial expansion, for instance). In recognition of this fact, some philosophers (for instance in the liberal tradition) have attempted to do justice to the 'fact of pluralism', and to eschew substantive conceptions of human nature and flourishing by providing 'formal' or procedural bases for normative claims justifying particular political arrangements. However, Butler argues that even such minimal accounts – for instance the Kantian-inspired conceptions of deliberation favoured by Rawls and Habermas – fail to purge themselves of unacceptably parochial assumptions: "Although the procedural method purports to make no substantive claims about what human beings are, it does implicitly call upon a certain rational capacity, and attributes to that rational capacity an inherent relation to universalizability."[1]

Does this mean that Butler eschews the project of providing a substantive, universal account of 'the human' that might support a normative political project? When

[1] Butler (2000: 15). Compare here my criticisms of Kant's focus on rationality in Chapter One.

discussing a structurally similar problem regarding the elaboration of an inclusive definition of 'women' – the ostensible 'subject of feminism' – Butler does eschew the project of providing such an account, opting instead for a politics of "open coalition [that] will affirm identities that are alternately instituted and relinquished according to the purposes at hand" (*GT* 22). While this model certainly retains its appeal for Butler in their later work, they nonetheless leave room for the possibility of a universal account of the human. While acknowledging the potential violence of false universals, Butler argues that "this does not mean that universality is by definition violent. It is not. But there are conditions under which it can exercise violence."[2] As we shall see, I think Butler does advance general claims about necessary features and conditions of embodied human life, and that their account has a genuine claim to universality. Furthermore, Butler thinks that their account of these features can do real normative work, and indeed can provide support for an egalitarian politics. I therefore think it would be premature to conclude that they absolutely disavow a substantive 'universalism'.

Nonetheless, they think the fact that (false) universality *can* exercise violence should give us pause. It certainly warrants a caution and humility in advancing ostensibly universal claims. It also shows the necessity of a critical project that identifies and exposes the exclusionary norms that serve to tacitly rationalise certain forms of violence. Such norms are not only to be found in philosophical anthropology and indeed in most cases the latter will only be thematising or providing ideological support for norms that are already implicit in certain political practices. Butler writes that

[2] Butler (2005: 7).

"sometimes the very terms that confer 'humanness' on some individuals are those that deprive certain other individuals of the possibility of achieving that status, producing a differential between the human and the less-than-human."[3] Their argument is that by demarcating the boundaries of 'the human', producing a standard by which humanity is measured, such terms can also produce another category – the 'less-than-human' – that forms the outside of that boundary. Those who do not 'fit' the dominate norms by which 'humanness' is conferred are consigned to that category.

What kinds of term does Butler have in mind? In an early version of this argument schema, they emphasise the role gender – or rather, the binary gender system predicated on a unity between gender, sex, and desire/sexuality – plays in the 'humanisation' of subjects: "Those bodily figures who do not fit into either gender fall outside the human, indeed, constitute the domain of the dehumanized and the abject against which the human itself is constituted" (*GT* 151). If one's *gender* is illegible or incoherent according to dominant norms – whether that is because one's 'sex' does not align with one's 'gender', or because one's desires are not oriented towards the gender 'opposite' to one's own in the assumed binary, or because one's sex or one's gender cannot be neatly situated in the terms of that binary – then one is at risk of not being recognised or 'read' as fully human, but rather seen as aberrant, abject, and even threatening and monstrous. The latter response, on Butler's view, is a reflection of the fact that the very illegibility of the abject really can threaten to undo, subvert, or undermine the dominant norm, in part by exposing its contingency. As they put it in a later text, this liminal figure is "a spectre that gnaws at the norms of recognition, an

[3] Butler (2004b: 2).

intensified figure vacillating as its inside and its outside. [...] As inside, it must be expelled to purify the norm; as outside, it threatens to undo the boundaries that limn the self" (*FW* 12). Since the human, Butler suggests, is "always already gendered" (*FW* 12), the appearance of a gender that cannot be 'read' also brings into question the stability of the norms by which we recognise the human.

Butler's formulation in the cited passage stands in need of modification. Firstly, it is not clear that having a 'coherent' gender is *sufficient* to guarantee 'humanisation' – even *within* the binary gender system, women have been cast as 'the abject'.[4] Plausibly, at least some 'exclusionary conceptions of the normatively human' have themselves been gendered (that is, masculinist). Secondly, it is not *exclusively* via gender that we are 'humanised'. It would not be correct to say that the 'illegibly' gendered subjects in question 'constitute the domain of the dehumanized' if that is understood to mean that they fully or exhaustively constitute that domain. Rather, as Butler registers explicitly in later writings, "the human is understood differentially depending on its race, the legibility of that race, its morphology, the recognizability of that morphology, its sex, the perceptual verifiability of that sex, its ethnicity, the categorical understanding of that ethnicity,"[5] and no doubt along more axes besides. In work written during the Bush era, their critical focus is primarily on the "ethnic frame" that is tacitly employed to justify both the suspension of human rights for detainees at Abu Ghraib and Guantanamo, and the disproportionate violence wrought by US military action in the Middle East in the wake of 9/11 (*PL* 16 and *passim*).

[4] Butler takes the notion of the abject from the work of Julia Kristeva; see e.g. Kristeva (2002a) and (2002b).
[5] Butler (2004: 2).

Returning to the passage from *Gender Trouble*, the phrase 'against which' is significant for Butler – the implication is that in order to be a meaningful category, the human must be constituted in contrast to something else. There is a sense in which this is an obvious truism: to talk of 'the human' implies a distinction from that which is not human. But we should not infer from this that the relevant contrast must always be 'the domain of the dehumanized and the abject'. In other words, the fact that there are exclusionary conceptions of the human does not mean that there can be no fully inclusionary ones. The positive task might be to draw the contrast in the right place, and for the right reasons.

However, Butler also highlights a negative and critical task here: to expose the norms that are exclusionary, that draw the contrast in the *wrong* place, and in doing so cast certain 'bodily figures' outside the boundaries of recognition. The use of the term 'bodily figures' here is also telling and important: if the category of the human is, on Butler's view, *constituted* by the exclusion of certain figures, then these figures, by definition, cannot be human – at least, according to the dominant conception of the human. So it is not accidental that Butler does *not* write here what it would be extremely natural to write: namely, that *human beings* who do not fit into either gender fall outside the human.

Butler thus wants to claim both that who we can recognise as human is a product of various constitutive but usually tacit norms of recognition, and that we can sometimes contest these norms on the basis of their exclusion of certain 'bodily figures' who

constitute their outside. Such lives, that are not quite *recognised* as fully living, can nonetheless be acknowledged, and such acknowledgement "can become the basis for a critique of norms of recognition" (*FW* 5). In order to explain how this acknowledgement might be possible, I turn now to Butler's ontological account of embodiment, which they believe can support their critical project.

Beginning with work written in the wake of 9/11 and in response to the Iraq War, Butler has been developing an ontology of the body that they hope will underpin a renewed commitment to egalitarian social conditions. This ontology understands embodied life in terms of its constitutive vulnerability and interdependency. Butler's ontology is meant to have a substantive normative upshot: it follows from recognition of our shared bodily condition, they argue, that we must reject violence and contest inegalitarian distributions of precarity in our social world.

However, the route from Butler's metaphysical claims to their ethico-political commitments is not always clear. Why should recognising our vulnerability prompt us to contest existing differentials of precarity and violence? After all, there can be no direct contradiction between an ontological condition of embodiment and particular social arrangements. Social arrangements, whatever form they take – whether just or unjust, egalitarian or inegalitarian, violent or nonviolent – will all be made possible by our shared ontological condition. Thus, it is not obvious from the point of view of *ontology* why any particular social arrangement should be normatively privileged over any other. Ontology, one might think, is simply neutral on the question of politics. Indeed, this objection has been raised by Butler's critics and commentators alike. Seyla

Benhabib argues that "social ontology, even one that is as sophisticated and psychoanalytically-inspired as is Butler's, [. . .] cannot lead us to normativity",[6] while Ann Murphy questions the legitimacy of a "scenario in which ontology is given priority [over ethics], or in which the 'ought' is somehow derived from the 'is'".[7]

In the remainder of this chapter, I aim to reconstruct Butler's response to this challenge. I begin in Section II by introducing Butler's 'new bodily ontology,' situating it in relation to 'existentialist' attempts to provide an ontological analysis of the human being. In Section III, I outline the key elements of Butler's ontology, explaining their conceptions of *interdependency*, *vulnerability*, and *precariousness*. I use Heidegger's distinction between the ontological and the ontic to illuminate the different senses these terms have for Butler. In the final three sections, I examine the relationship between Butler's ontology and their normative claims. I argue in Section IV that one of their central attempts to derive an ethic of nonviolence directly from our interdependency is unpersuasive, resting on a conflation of that term's ontic and ontological senses. Nonetheless, I argue in Section V that Butler is right to suggest that genuinely acknowledging our vulnerability is likely to make us more responsive to the claim of the other, and to loosen the grip of the 'military fantasies' encouraged by an idea of absolute sovereignty. A commitment to nonviolence and a more egalitarian shared human condition amounts to an ontic mode of *acknowledging* our ontological condition, while the illusory ideal of sovereign independence and the violence it encourages amount to a *disavowal* of it. In Section VI, I conclude by briefly addressing

[6] Benhabib (2013: 152).
[7] Murphy (2011: 588). Admittedly, Murphy argues that Butler does not in fact give ontology priority, despite the passages in their work suggesting this. See fn. 59 below for critical discussion of Murphy's account.

the question of what makes these particular ways of relating to our ontological condition good or bad. I do so by examining two political stances – individualism and a narrow 'politics of vulnerability' – and showing why these are politically inadequate from Butler's perspective.

It is worth repeating that my aim is a reconstructive one. Butler does not always draw a clear distinction between the ontological and ontic senses of their key terms that I identify. For this reason, my formulations conflict somewhat with the letter of Butler's explanation of their bodily and social ontology. Nonetheless, I hope the usefulness of framing their project in these terms will be demonstrated in my discussion of the specific ontological features of embodiment to which Butler draws attention. Furthermore, whether or not one accepts the ontological or political views I attribute to Butler, my broader aim is to use their work to show that there is an interesting and plausible path from the ontological to the normative. Thus, my general conclusions in this chapter will be of interest even to those unsympathetic to the specific claims Butler advances.

II. Butler's 'New Bodily Ontology'

Butler introduces their project in *Frames of War* as follows:

> I want to argue that if we are to make broader social and political claims about rights of protection and entitlements to persistence and flourishing, we will first have to be supported by a new bodily ontology, one that implies the rethinking of precariousness, vulnerability, injurability, interdependency, exposure, bodily persistence, desire, work and the claims of language and social belonging. (*FW* 2)

From the rather formidable list that closes this passage, it is clear that providing a new bodily ontology will involve 'rethinking' a whole constellation of concepts. In the next section I will give a brief outline of Butler's conception of three of these: interdependency, vulnerability, and precariousness. Since the concepts making up Butler's constellation are closely interconnected, however, the discussion will also shed light on their understanding of injurability, exposure, and desire.

First, allow me to make some general observations about the aims and scope of Butler's project. To begin with, the ontological claims Butler makes are meant to be a prerequisite for 'broader' normative claims, in at least two senses. First, their ontological claims are meant to prompt an extension of rights and entitlements to human beings and populations currently excluded from them; the sphere of ethico-political concern is thereby broadened. Second, their ontology is supposed to prompt us to rethink what would be involved in securing those rights and entitlements. Butler's view is that understanding these only in terms of the rights *of individuals* is insufficient. We must also take into consideration and ensure the protection of social *ties* between individuals, and the social *conditions* for flourishing. I shall return to these points.

Butler is not pursuing ontology in the sense most prominent in analytic philosophy, namely the project of determining what kinds of entities there are.[8] Rather, Butler's 'bodily ontology' is an account of what it is to *be* a body. Butler seeks to do justice to the fact that a body is never completely independent of other bodies, nor of social

[8] The 'analytic' conception of ontology is strongly indebted to Quine, especially the first essay in Quine (1980). As we shall see, I take Butler's approach to ontology to be closer to that of early Heidegger than to that of Quine. I am grateful to Nikhil Krishnan for pressing me to be more explicit about the conception of 'ontology' in play here.

norms that constitute what a recognisable body is. The notions of sociality and embodiment are interdependent for Butler: to be a body is always already to be exposed to "socially and politically articulated forces as well as to claims of sociality [. . .] that make possible the body's persisting and flourishing" (*FW* 3). It is misleading to treat a body in complete isolation from its connections to social forces and claims. Part of Butler's target, then, is the "individualism" that is often presupposed in contemporary politics and philosophy, and its attendant conception of personhood (*FW* 5, 20, 145).[9]

Furthermore, the bodies in question are *living* bodies. Butler is not offering a general account of what it is to be a material thing. Of course, not all living bodies are human bodies. Butler's ontological claims will have normative consequences that extend to our treatment of non-human animals, and of the inanimate natural environment insofar as it is a shared support of living creatures.[10] Elaborating the conditions for human *life* involves showing what is needed to sustain both human and non-human animal life. Butler's account is therefore not, or not primarily, an *anthropology*, or an account of a specific difference between humans and other animals. Their ontological claims thus concern not so much our species-being as our 'genus-being'. Nevertheless, an account that gives ontological conditions for the possibility of life in general will *a fortiori* elucidate necessary conditions for human life in particular.[11] Embodiment is something that we share with other animals, and is indeed a constitutive feature of animality.[12]

---

[9] In *Frames of War*, Butler leaves the critique of individualism mostly implicit. For a discussion of the contribution of their ontology to such a critique, see *The Force of Nonviolence*, especially Chapter 1. See also Section VI below.

[10] For Butler, precariousness is both a "shared condition of human life," and "a condition that links human and non-human animals" (*FW* 13; see also xxv, xxx).

[11] One of these conditions will turn out to be vulnerability; see Section III. Like Butler, Erinn Gilson emphasises that vulnerability is part of the human *condition*, "in contradistinction to 'human nature'"(2011: 327, fn. 5), but can also be understood as a condition of non-human

Butler takes pains to distance their view from transhistorical accounts that downplay the contribution made by social factors to the constitution of bodies:

> To refer to 'ontology' in this regard is not to lay claim to a description of fundamental structures of being that are distinct from any and all social and political organization. [. . .] It is not possible first to define the ontology of the body and then to refer to the social significations the body assumes. Rather, to be a body is to be exposed to social crafting and form, and that is what makes the ontology of the body a social ontology. (*FW* 2-3)

The phrase 'fundamental structures of being' might put us in mind of Heidegger's attempt in *Being and Time* to provide an 'existential analytic of Dasein' – to uncover the necessary ontological structures, the 'existentialia', that underlie any *particular* way in which 'we ourselves' can be.[13] Heidegger introduces a distinction between the *ontological* and the *ontic* (*BT* 31/H11). Ontology is the inquiry into *Being*, "that which determines entities as entities, that on the basis of which entities are already understood" (*BT* 25-26/H6), whereas ontic or ontical sciences deal exclusively with entities themselves. Philosophy is ontological, and Heidegger is concerned to distance it from the ontic; the existential analytic is distinguished from biology, psychology, and anthropology, for

---

animal life and perhaps nature in general. Gilson's further claim that we have no grounds for understanding vulnerability as "an innate feature of human beings" (2011: 327), however, seems overly hasty. Vulnerability in Butler's sense may not be a *unique* feature of human beings, but it is for all that an *innate* one, insofar as 'innate' means "belonging to the essential constitution" (*OED*).

[12] Butler's account thus seeks to "take seriously the chiasmic relation implied by the phrase, 'the human animal'" (*FW* 17; see also 7). It is unclear whether Butler's metaphor is grammatical ('chiasmus') or biological ('chiasma'), but in either case the suggestion is of an 'intercrossing.' The point seems to be that our humanity cannot be extricated from our animality (and thus our embodiment). Interestingly, Adrian Moore has argued that Kant attempted to break from the Aristotelian tradition and do precisely that (2003: 138, 79).

[13] See *BT* 36/H15 and 75/H50 for the identification of Dasein as 'we ourselves,' and *BT* 71/H45 and *passim* for 'existentialia' (and its singular form, 'existentiale'). This kind of project is also pursued by other existential phenomenologists, such as Jean-Paul Sartre; see Sartre (2003).

example (*BT* §10).[14] Although Butler does not invoke the distinction between the ontological and the ontic explicitly, I will seek to show its usefulness in clarifying their position, and I shall refer to it frequently.[15] While Butler sometimes seems to utilize something like this distinction, at other times their failure to draw it explicitly results in problematic equivocations.[16]

The distinction between the ontological and the ontic gives us a way of positioning Butler in a critical relation to the project of existential ontology. In seeking the *a priori* structures of existence, Heidegger has to place social and political organization on the side of the 'ontic' – these are merely 'factical' or 'existentiell' arrangements of society that Dasein might find itself in, made possible by the more basic structures of existence: being-in-the-world, being-with others, &c. These latter can be described without reference to any *particular* social arrangement. Butler, on the other hand, takes ontology itself to be partly constituted by historically contingent social arrangements and power relations, citing Foucault's notion of the "historical *a priori*" in this connection (*FW* 6).[17] Placing concrete social relations in the realm of the 'ontic' thus obscures the fact

[14] See Chapter Two, Section I above for discussion, especially fn. 8.
[15] Aret Karademir also "read[s] Butler with the aid of Heideggerian concepts" from *Being and Time* (2014: 825), although he does not utilize the ontological/ontic distinction in his reading. His focus is on Butler's early work on gender and its relation to freedom, rather than on their later, more avowedly ontological work. Karademir views Butler's "social constructivism" as distancing their work and Heidegger's, which "interrogates the universal structures of human existence" (2014: 824). As will become clear, my suggestion is that some of the central claims of Butler's later work can fruitfully be read as contributing to precisely this latter kind of 'interrogation', though the ontological structures Butler identifies differ markedly from those analyzed by Heidegger.
[16] See Section IV.
[17] Sacha Golob has suggested that the epochs of Heidegger's history of Being can also be understood in terms of this Foucauldian notion; see Golob (2017: 630). Thus, although my aim later in this chapter will be to draw on Heidegger's *early* work to shed light on aspects of Butler's ontological project, there may also be parallels between Butler's project and

that what a human being is, what a subject is, indeed what a *body* is, is not entirely historically stable – it obscures the constitutive role of discourse and its material effects.[18]

The account given so far is an oversimplification, in that it is not obvious that the 'existentialists'[19] really thought ontology could dispense with discussion of 'any and all social and political organization'.[20] The more important point to consider here, however, is that it is not obvious that Butler's *own* account can avoid laying claim to a description of very general 'fundamental structures of [embodied] being'. There seems to be a non-sequitur in the passage quoted above: we are told that 'it is not possible *first* to define the ontology of the body and *then* to refer to the social significations the body assumes', and this negative claim is apparently justified by the fact that 'to be a body is to be exposed to social crafting and form'. But this latter claim is itself extremely general – indeed, it is plausibly read as elucidating a 'fundamental structure' of the

---

Heidegger's later, more overtly historical approach. I am grateful to Mark Wrathall for emphasising this latter point to me.

[18] The 'entirely' poses difficulties. See Srinivasan (2018), §2 for a discussion of the (ineffable?) body that exceeds language on Butler's account.

[19] I use 'existentialists' and 'existentialism' in scare quotes to group Heidegger with Sartre and Beauvoir. This grouping is convenient, but somewhat tendentious; although Heidegger's early work inaugurated the tradition of existential phenomenology, he later disavowed the existentialist appropriation of his work. See his comments on Sartrean existentialism in Heidegger (*LH* 250-51, 254). When I use the terms *without* scare quotes, I intend to exclude Heidegger.

[20] In different ways, Heidegger and Sartre give a prominent role to sociality. For Heidegger, there can be no Dasein without *Mitsein*, 'being-with' (*BT* §§25-26); we have an irreducible relation to others. Heidegger's discussion of historicality and the 'heritage' of possibilities we inherit (*BT* §74, especially 435/H383-84; see also 167/H129) also suggests that he views our understanding of the world and ourselves as socially mediated. Sartre, for his part, dedicates the entire third part of *Being and Nothingness* to discussing relations with Others. Finally, while neither of these philosophers devotes much space to discussion of concrete political organization in these works, this certainly could not be said of Simone de Beauvoir, another important existentialist philosopher. I consider Beauvoir's conception of the relation between ontology and politics in more detail in Section V, arguing that her position bears affinity to Butler's.

being of the body, *any* body. The particular kinds and effects of social and 'discursive' crafting by which the body is constituted and to which it is exposed are historically variant; what is *not* historically variant is the 'fact' of exposure to social crafting.[21]

This suggests that, for Butler, ontology itself can operate on two levels. The first is the level of 'fundamental structures of being'. As we shall see, Butler's account does attempt to elucidate necessary, transhistorical ontological features of embodiment: part of what it is to be a living body is to be vulnerable, exposed to social crafting, dependent, &c. The account at this level will have to do justice to sociality, insofar as our account of embodied being had better acknowledge the fact that an 'individual' body is always already a 'social' body, subject to all manner of impingement by social forces 'outside' of it. The second level would provide a 'social ontology' in a different sense, describing the norms that constitute the being of specifically social phenomena, such as political categories (e.g. 'citizen'), gender categories, and what one might call 'normative ideals of the body': the contingent, historically variable social and discursive norms that regulate what can be recognized as a (fully) human body.[22] This second level is still 'ontological' – it still concerns what it is to *be* a citizen, a woman, an intelligible body. But it is a *social* ontology because a full account of the being of such entities must make reference to social norms local to their historical context.

[21] One could arguably draw a similar distinction regarding Butler's early account of gender performativity: what is not historically variant is that gender is performatively constituted, but the particular 'legible' formations of gender that are produced through its performance *are* historically variant. See *GT* Chapter 3, Section 4, and the Conclusion, for the classic statement of Butler's view; see also Butler (1988).

[22] Butler's discussion of "contingent ontologies" of gender (*GT* 45–46) is strongly suggestive of this 'second level' of ontology; their later work seems to operate at both levels.

Butler is interested in pursuing ontology at both of these levels, though I shall focus here on the first level – on general structural features of embodiment. But Butler is also interested in particular *ontic* situations that are made possible by these ontological features, and in the differential distribution of certain politically important ontic properties. We are all equally 'vulnerable' in a general ontological sense, but we are *not* all equally 'vulnerable' in the ontic sense. Furthermore, Butler's critique of ontic violence and inequality is meant to proceed via an appeal to our ontological condition. As we shall see, apprehending a *general* condition of (ontological) 'precariousness' is meant to lead us to contest unequal and differential distributions of (ontic) 'precarity'.

III. A Post-Existential Analytic? (Inter)dependency, Vulnerability, Precariousness

Three ontological features of embodiment are especially central to Butler's project: interdependency, vulnerability, and precariousness. These features play an analogous role in Butler's account of embodiment to that played by the existentialia in Heidegger's analysis of existence. Butler understands them to be constitutive features or conditions of embodied life. In this section, I examine them in turn, showing how a distinction between an 'ontological' employment and related 'ontic' uses of each term is in play in Butler's discussion.[23]

(Inter)dependency:

Ordinarily, we think of the passage from dependency to independence as closely

[23] My account abstracts somewhat from changes in emphasis across Butler's work. The arguments of Butler's latest book, *The Force of Nonviolence*, draw particularly on interdependency, downplaying to some extent the significance of precariousness, which appeared as the "central concept" of Butler's ontology in earlier works; see Gilson (2014: 43). Nonetheless, Butler continues to view these phenomena as deeply interconnected. The apparent change in emphasis is therefore not indicative of a fundamental change of view.

connected with the passage from infancy to childhood to adulthood. As infants, we are utterly dependent on the care of others; as Butler puts it, "none of us come into the world as self-sufficient beings".[24] Our passage to 'maturity' amounts to an overcoming of this dependency, and 'independence' thus acquires a positive valence. However, on Butler's view, we *never* become fully self-sufficient, and we are always dependent on others in various ways, and they on us: the "struggle of dependency and separation [. . .] does not merely designate a stage of childhood to be surmounted" (*FW* 183). Simply in virtue of having needs that may or may not be met, we are always materially dependent on others and on a supportive environment. Butler relates this condition closely to our embodiment, going so far as to characterize "the human body [itself] as a certain kind of *dependency* on infrastructure, understood complexly as environment, social relations, and networks of support and sustenance" (*RVR* 21). Absolute self-sufficiency, absolute independence, is therefore a fantasy, and relative self-sufficiency is far from being guaranteed:

> Those of us who achieve some measure of self-sufficiency do so only be relying on social structures, including material infrastructures, that have been secured through political and economic means. Even our ability to stand, if we have that ability, depends on an existing surface or platform that provides the possibility of achieving balance and mobility.[25]

The example of the ability to stand suggests two different senses of 'dependency'. We can talk both of specific *ontic* dependencies, and a more fundamental *ontological* dependency. Which ontic dependencies characterize a particular person's life will be a matter of their *situation* (to appropriate another piece of 'existentialist' vocabulary).

[24] Butler (2018: 246).
[25] Butler (2018: 246-7).

Whether I have the ability to stand or not makes a difference regarding how social spaces must be organized to be accessible or navigable for me. I thus depend on my environment and on others in various concrete ontic ways. But there is a deeper, ontological sense of dependency that underlies these. To *stand* I must be supported by some particular surface or other; to *be embodied* is to be dependent at all times on infrastructure in Butler's broad sense, on an entire social and material world. This dependency is thus a necessary, ontological feature of our embodiment.

Butler draws a further consequence: our condition of dependency is reciprocal, and is therefore a condition of *inter*dependency. We depend on others, and they on us; we all depend on a shared social world, which we in turn help to sustain – or destroy. This interdependency is partly material, but it is also a precondition of identity and subjectivity on Butler's view. This is "finally a Hegelian point" (*FW* 43), both recalling and transforming Hegel's idea that self-consciousness is dependent on the recognition of an Other.[26] One of Butler's central arguments for 'nonviolence' will hinge on this claim: their view is that in 'undoing' the Other, one undoes oneself.[27]

Vulnerability:

It would be natural to construe 'vulnerability' as a special susceptibility to injury (physical or otherwise), to being wounded or subjected to violence. In positing an ontological condition of vulnerability, Butler certainly has this sense in mind. Connecting the three notions presently under discussion, they write: "To be dependent implies vulnerability: one is vulnerable to the social structure upon which one depends,

[26] See Hegel (2018), §§178-96.
[27] I discuss this argument critically in the next section.

so if the structure fails, one is exposed to a precarious condition" (*FN* 46). Here, the implication is that *because* we always need and depend on a supporting social structure, we are always vulnerable to its possible withdrawal or collapse, always potentially exposed to injury and precarity. However, Butler is careful to resist any straightforward reduction of vulnerability to injurability (*FW* 34, 61). Their suggestion is that there is a broad sense of vulnerability that makes 'injurability' possible. Injurability is only one modality of vulnerability; being injurable is what Heidegger might have called a 'founded mode' of being vulnerable (cf. *BT* §13). As Butler puts it, "the body that exists in its exposure and proximity to others, to external force, to all that might subjugate and subdue it, is vulnerable to injury; injury is the *exploitation* of that vulnerability" (*FW* 61, emphasis added). This fundamental exposure is what Butler calls a "primary vulnerability to others" (*PL* xiv), a "primary human vulnerability" (*PL* 28). In the terms of the ontological/ontic distinction, this primary vulnerability, this fundamental 'exposure' to others, is *ontological*. The more specific modes of vulnerability that it makes possible are *ontic*.[28]

One such mode is vulnerability to injury or harm; some human beings may be more vulnerable to injury, or certain kinds of injury, than others. However, this is far from

[28] It is often recognized in the literature on Butler that they have a two level or 'dualistic account' of vulnerability: see Lloyd (2015: 185, cf. 175) and Gilson (2011: 310); cf. Murphy (2011: 578). Commentators generally employ a contrast between 'primary vulnerability' and concrete forms of vulnerability, rather than couching the discussion in terms of the ontological/ontic distinction I develop here. Moya Lloyd comes close to marking this distinction when she notes that (primary) vulnerability is a "*condition of possibility* for love, desire, care, hope and life" (2015: 172, emphasis added). Gilson also notes that "vulner*ability*" is a "*condition* of potential that *makes possible* other conditions" (2011: 310, some emphasis added), including 'positive' modalities of vulnerability such as those mentioned by Lloyd, which I discuss further in the main text. Gilson emphasizes that we ought "to refuse to conflate vulnerability in its most profound and general sense [. . .] with specifically negative forms of vulnerability" (2011: 324). I discuss some consequences of such a conflation in Section VI below.

being the only important ontic mode of vulnerability on Butler's conception. On the ontic level, as Ann Murphy puts it, "no body is vulnerable in exactly the same way as any other" (2011: 578). There are various ways in which we can be vulnerable, and the fact that bodily vulnerability can be exploited in violence nevertheless leaves opens the possibility of a *non*-exploitative relation to others' vulnerability. What is *definitive* of primary or ontological vulnerability is not the possibility of its exploitation (that is, injurability), but rather the more general condition of 'exposure and proximity to others' – to the "obtrusive alterity" the body "comes up against" (*FW* 34). This obtrusion is precisely what "animates responsiveness" to the world, where responsiveness includes "a wide range of affects: pleasure, rage, suffering, hope, to name a few" (*FW* 34). Butler's notion of vulnerability is, then, a markedly ambivalent one. The ontological condition that underlies our injurability ('suffering') also underlies affects with a more positive valence ('pleasure', 'hope'), or whose valence is itself ambiguous ('rage', which Butler thinks can play an important role in critique and political resistance). Vulnerability, the condition of being necessarily "exposed to others" (*FW* 33) is not in itself something bad, something we should want to deny or will away.

While this notion of vulnerability expands on the ordinary use, then, it is not discontinuous with it. Nor would Butler want to deny that we should keep using 'vulnerable' in a narrower sense. Doing so can be a way of registering injustice: registering the fact that in violence, vulnerability can be differentially exploited.[29] But *ontological* vulnerability is a condition of possibility of *various* forms of ordinary 'ontic'

[29] Nonetheless, Butler has reservations about a 'politics of vulnerability' grounded on the identification of particular "vulnerable populations" (*FN* 186-92). See Section VI below.

vulnerability, and Butler's notion shows how vulnerability in a narrow and negative sense is necessarily bound up with more positive modalities of our exposure to others. Without a primary vulnerability to others, there would be no fear of injury, but also no possibility of affection, pleasure, or intimacy. Part of the task of ethics is to minimize the 'negative' modalities of vulnerability, especially where their distribution is inegalitarian, and to cultivate its 'positive' modalities.

Precariousness and Precarity:

Butler's discussion of precariousness and precarity also involves distinguishing between a general ontological condition, and specific, differential ontic modalities: "the more or less existential conception of 'precariousness' is thus linked with a more specifically political notion of 'precarity'" (*FW* 3). The link between the two is structurally similar to the relation between the two senses of 'vulnerable' identified above: the former (ontological, 'more or less existential') is a *condition of possibility* of the latter (ontic, 'specifically political', capable of differential distribution).[30] How should we understand each term? Butler glosses 'precariousness' as the fact "that life requires various social and economic conditions to be met in order to be sustained as a life" (*FW* 14). This fact is a completely general one, and is, as it were, indifferent to the political, 'ontic' matter of whether those social and economic conditions are in fact met for a given life or

[30] Catherine Mills uses the terms "ontological and ontic" in passing to describe the "differentiation of different kinds of precariousness" (2015: 45), though she does not develop a reading that uses the ontological/ontic distinction *systematically* to clarify Butler's ontology, as I attempt to here. The 'different kinds' Mills is referring to map onto the precariousness/precarity distinction, which she goes on to discuss. She rightly points out that Butler does not always maintain this distinction, occasionally conflating the two senses of 'precarious' (2015: 45-9). As I argue in Section IV, a similar conflation afflicts Butler's discussion of interdependency.

population. Hence, Butler talks of a "generalized precariousness" and a "generalized condition of precariousness". Indeed, "lives are by definition precarious" (*FW* 25).

'Precarity', on the other hand, "designates that politically induced condition in which certain populations suffer from failing social and economic networks of support and become differentially exposed to injury, violence, and death" (*FW* 25). It is thus an 'ontic' condition or state: particular populations are especially precarious in this latter sense.[31] This state is 'politically induced', the result of both political and social formations. Indeed, Butler goes further, suggesting that the "differential distribution of precarity is at once a material and a perceptual issue" (*FW* 25). On Butler's view, social norms structure perception, delimiting who can be perceived as fully human, and the distribution of rights and material goods is thus partly determined by the field of possible perception.

Butler provides a powerful 'case study' of this idea in their early essay 'Endangered/Endangering: Schematic Racism and White Paranoia'.[32] Butler argues that the acquittal of police officers who were tried for the use of excessive force against Rodney King, an unarmed black man, could only be possible on the basis of a prior

[31] I would therefore contest Gilson's claim that "precarity is just precariousness exacerbated" (2014: 45). Gilson does go on to note that precariousness (unlike precarity) is a general condition, but her gloss here fails to capture the sense in which Butler is talking about two different levels (ontological on the one hand, and ontic and political on the other). Similarly, Gilson writes that "whereas increased precariousness produces increased exposure to 'injury, violence, and death[,]' the consequences of increased vulnerability are indeterminate" (2014: 47). To talk of an 'increase' here strikes me as infelicitous if we remain at the ontological level: our condition of ontological precariousness and vulnerability is, as it were, a constant. We can talk of an increase in *precarity*, or in particular *modes* of vulnerability. But the reason the consequences of increased vulnerability might appear 'indeterminate' here is that the particular mode of *ontic* vulnerability in question is left undetermined. (These remarks apply, *mutatis mutandis*, to 'dependency', as we shall see in the next section.)

[32] I am grateful to an anonymous referee at *Ergo* for suggesting that I address this paper here.

"racist production of the visual field".[33] Jury members were shown a video recording that, as Butler unhesitatingly (and rightly) puts it, "shows a man being brutally beaten, repeatedly, and without visible resistance", but one juror claimed that "she believed Rodney King was in 'total control' of the situation" – a perverse "feat of interpretation" apparently typical of the jury, given the eventual verdict.[34] Butler argues that such an interpretation could only arise on the basis of a prior framing and schematization of the visual field, according to which "the black male body is always already performing [a threatening action] within that white racist imaginary".[35] Thus, a hand raised in self-defence to protect a black man from the blows of the police batons is immediately 'read' as signifying and foreshadowing a violent intent.[36] In *Precarious Life*, Butler gives a somewhat similar analysis of the racist "bestialization" of prisoners detained indefinitely in Guantanamo Bay, who are figured as *essentially* exceptional individuals who "must be constrained in order not to kill, [who] are effectively reducible to a desire to kill" (*PL* 78). This figuration is deployed in order to legitimate the suspension of international codes guaranteeing these people's rights.

Butler's point, then, is that the structuring of the field of perception can underlie and legitimize certain forms of violence, which then have a knock-on effect on the forms of precarity (and indeed vulnerability) to which certain populations are disproportionately exposed. A critical analysis of these perceptual schematizations is thus a necessary element of an explanation of differential precarity, though Butler does not claim that it

[33] Butler (1993: 22).
[34] Butler (1993: 15).
[35] Butler (1993: 19).
[36] Butler (1993: 20).

is a sufficient explanation in itself.[37] This suggests that part of the explanation for this differential distribution will be given at the second level of ontology identified in Section II: the level of constitutive but socially specific norms. By exposing the norms that constitute who can be recognized as fully human, we can understand how those norms support particular distributions of precarity.[38]

IV. From Ontology to Politics: Interdependency and Nonviolence

What normative consequences can be drawn from this ontological account of embodiment? Butler attempts to move directly to an ethic of nonviolence from the interdependency that is a constitutive condition of embodied life. Consider the following two passages:

> We can assert in a general way that social interdependency characterizes life, and then proceed to account for violence as an attack on that interdependency, an attack on persons, yes; but perhaps most fundamentally, it is an attack on "bonds." (*FN* 16)

> Violence against the other is [. . .] violence against oneself, something that becomes clear when we recognize that violence assaults the living interdependency that is, or should be, our social world. (*FN* 25)

[37] Lloyd suggests that Butler rarely, if ever, "explore[s] in detail the actual mechanisms that give rise to the concrete precaritisation of a particular population beyond referring to it as a general 'political' process" (2015: 176). As I have suggested, the schematization of the perceptual field can be viewed as one such 'mechanism' of precaritisation, though not the only one. More broadly, however, the potential fruitfulness of Butler's account perhaps lies in its provision of a framework for discussing *various* modes of precarity and mechanisms of its production, regardless of Butler's own success in explaining particular cases.

[38] The question of *how* we can come to recognize and contest such norms, given that they structure our perception at a very basic level, is an important one for Butler. Lloyd worries that "it is not clear how *in determinate conditions* it is possible, if at all, to overcome the normative constrictions that prevent us receiving an ethical address" from those who are not perceived as fully human (2015: 186). I cannot pursue this question further here, though I hope to address it in future work.

Butler's point is not just the familiar idea that 'violence begets violence'. Rather, their claim is that there is something self-undermining about acts of violence against the 'bonds' that partly constitute the agent of that violence. A violent act against the other is *directly* violence against oneself, because one's 'self' is dependent upon its bond to the other.

It is not entirely clear whom Butler seeks to address with this argument. While some of the arguments Butler pursues for nonviolence address a presumptively left-wing audience, dealing with the question of whether tactical violence can be justified as a means to progressive or liberatory ends (*FN* 12-3), the appeal to the consequence of 'violence against oneself' here suggests that the argument is intended to have a more general import. Butler is interested in acts of self-defence as exceptions to prohibitions on violence. The intuitive, philosophically widespread view is that violence is justified (only) if it is committed in self-defence. But Butler responds that the 'self' to be defended in such cases cannot be straightforwardly delineated from the other. If the argument for nonviolence is successful in such a case, it also promises to speak more generally. Perhaps the argument is meant to have force even for a more or less egoistic agent, who does not yet acknowledge the claims of others. If the egoist recognizes their dependence on others, and further recognizes that violence is an assault on interdependency, on social bonds that sustain the self, they thereby recognize that they have a reason not to commit violence; the violence they exact on others affects them as well.

If this is the intention behind the argument, it is surely unsuccessful. For it trades on an ambiguity concerning the ontological and ontic senses of 'dependency' identified above. The argument seems to invoke the idea of an assault on interdependency *as such* – not this or that particular relation of dependence, this or that social 'tie', but the whole social world. As Butler claims elsewhere, a self that sought self-preservation through violence "stands worldless, threatening this world" (*FN* 149). This might lead us towards an 'ontological' interpretation of the term. But does it make sense to talk about an assault on our interdependency in *that* sense? What could it mean for a self to stand genuinely 'worldless', rather than standing in an impoverished or cloistered world?

If interdependency is part of what it is to be a living being, then an assault on 'living interdependency' *as such* would amount to a threat to our very being. The 'destruction' of our interdependency would mean *our* destruction. There may indeed be cases we would describe in such terms: our continued violence against the environment threatens to make life on our planet literally unliveable; nuclear disasters might have similarly grim consequences. But these examples, urgent though they are, make the appeal to interdependency superfluous. These are cases where the self-destructive nature of our violence is, as it were, transparent. One need not recognize our social interdependency to recognize that such scenarios could utterly destroy the conditions for life. Of course, in the case of climate change such recognition has in any case proved insufficient to override the shortsighted profit motives of fossil fuel companies, to take just one example.[39] This is something Butler acknowledges when discussing the

[39] See Fraser (2021: 100-7).

environmental policy of countries such as the US: "perhaps they do know that they are in the midst of a globally destructive activity, and that too seems to them like a right [. . .] that should be compromised by nothing and no one" (*FN* 44).

On the other hand, if we understand Butler's argument as invoking the idea of an *ontic* interdependency, it matters a great deal *which* ontic social bonds are threatened by a specific act of violence. For why should an egoistic agent not be willing to accept *some* violence to themselves if compensated by some gain, especially if the dissolution of *some* social bond or other is unlikely to have a great impact on them? We are fundamentally (inter)dependent creatures. Perhaps the boundaries of the 'self' are unclear. But it does not follow that there are *no* such boundaries, or that the self could not survive the destruction of *some* of the social ties that it depends upon and that relate it to others.

The particular ontic dependencies each of us has are not symmetrical. The CEO of a transnational corporation depends on the workers in their warehouses, if their business is to be profitable. But the workers' entire livelihood is dependent on keeping their substandard wage. The workers have a certain fungibility from the perspective of the company: if one is fired, someone else will be willing to take on the work. As individuals, workers are not able to effectively contest their treatment (to negotiate their wage, or to raise safety concerns, for instance): hence the political significance of unionization. It is not clear in what sense the exploitation and violence enacted on the worker in this situation is also violence against the CEO. Further, even if there *is* any such sense, it is not clear that it would provide the CEO with any new and overriding motivation towards nonviolence. Indeed, as Amia Srinivasan has suggested, it may be

that some selves – including, plausibly, our CEO – depend precisely on violent exploitation to maintain themselves, their life, their ever-increasing 'worth' and social position.[40] If so, it is unclear how recognition of that dependency could motivate them to abandon their exploitative practices. From an egoistic perspective, perhaps the (prudentially) rational response to that recognition would be to shore up exploitation and ensure its perpetuation.

This leads to a more general worry about the sort of project Butler is pursuing, one that aims to make its way from ontology to normativity. The worry is this: ontology is normatively neutral. If we describe the *ontological* condition of interdependency or vulnerability, we are describing something that is in a sense inalienable. To live *just is* to be vulnerable and dependent. But if that is the case, ontology has nothing to say about the particular (ontic) modalities of dependency and vulnerability. No particular political arrangement can 'contradict' our ontological condition, so the latter places no constraint on the former.[41] Ontological dependency is not something that can be assaulted or destroyed, except in the sense that life itself can be destroyed. Its protection cannot be made into a political goal, and it offers no guidance regarding which political arrangements are to be preferred. Whatever ontic distribution of precarity, dependency, and vulnerability to violence obtains, our ontological condition remains untouched, for it is the background condition of *any* such arrangement. It is therefore tempting to be

[40] See her conversation with Butler: "Judith Butler: The Force of Nonviolence," https://www.youtube.com/watch?v=HN5D9rlkRcA (40:17–42:50).

[41] There is nonetheless an obvious way in which ontology *might* constrain normativity: ontology might limit what is *possible* or, more controversially, what is feasible. For a (professedly opinionated) introduction to issues concerning the relevance of feasibility to normative questions, see Southwood (2018). See especially Section 2.1 for a brief discussion of the idea that feasibility might place constraints on the normative domain. I shall leave this sort of constraint aside, as it is clearly not the sort of connection that interests Butler.

sceptical about the project of deriving normative claims from an ontology, whose proper domain is at best descriptive. Seyla Benhabib puts this point in stark terms:

> Ethics is not social ontology. Social ontology, even one that is as sophisticated and psychoanalytically-inspired as is Butler's, can help disclose the permutations of self-other relations as well as uncover the necessary bases for the formations of receptivity [. . .] but it cannot lead us to normativity.[42]

V. Vulnerability, Fantasy, and Acknowledgement

Butler's 'bodily ontology' might seem, to many, undeniable. Many will find it intuitively obvious that we are exposed to others; that our persistence and flourishing depend on conditions outside ourselves; that we are always dependent on others, and they on us. Is Butler's account then trivial? Did we not see Butler claim that it is true 'by definition' that life is precarious? In this section, I will address this worry about triviality. In doing so, I hope to show that Butler also has resources to respond to the problem identified at the end of the previous section – that there is a bridge from ontology to normativity.

If Butler's ontological claims appear trivial, it may be difficult to see what form opposition to those claims might take, and why anyone might be resistant to them. But Butler insists that there is such opposition "to vulnerability", and suggests that "it is probably best not to regard this opposition primarily as an 'argument'" (*PL* 19). To see what they have in mind here, consider Kate Manne's characterisation of Butler's position:

> We (a pronoun we can allow to have a somewhat open-ended referent) are simultaneously agents and subjects – acted upon as we act, and mutually

[42] Benhabib (2013: 152).

> impressionable via sense and sensibility. It is not just that others impinge on us causally, and make a difference to our sense of self, *which nobody would deny*. A *constitutive* dependency – or better, entanglement – is at issue here. Subjectivity is not thinkable without inter-subjectivity. Who I am depends partly on how I am regarded, treated, addressed, called upon, and spoken of by other subjects, with whom I share a historical, social, and material world.[43]

While there is much that is right about this characterization, I want to draw attention to the phrase 'which nobody would deny'. In the latter part of this passage, Manne places the emphasis on what could be called our 'discursive' interdependency, while downplaying the interest of Butler's commitment to our 'material' interdependency. It is the latter that Manne seems to have in mind when she claims that nobody would deny that others 'impinge on us causally'. There is of course a sense in which this *is* undeniable, as Butler is well aware. And yet, it is an important feature of Butler's account that we nonetheless *do* deny the very obvious fact of our material dependence on others, and the symmetrical facts of their material dependence on us, their vulnerability to us, and our constant 'causal impingement' on them.

The most important point to make in relation to Manne's claim is that there are *various* kinds of denial (or illusion, or fantasy) that might have different sorts of motivation and origin. Manne seems to have a highly intellectualistic conception of denial in mind: no one would deny these facts, because they are self-evident. But we can also talk about denial in a psychoanalytic register, for instance when we say that someone is 'in denial'. It is arguably a precondition of being in denial that on some level one knows precisely that which one is in denial about, so being in a position to know something just in virtue of reflecting on it ('self-evident') is perfectly consistent with denying it in

[43] Manne (2018: 234, first emphasis added).

this sense. Finally, and most importantly in this context, we can speak of denial in the sense of *disavowal*. 'Disavowal' has greater normative force than 'in denial'. It is not just an epistemic notion, but an overtly moral one. To disavow something is to refuse to associate oneself with it, to refuse to acknowledge it – to disclaim knowledge of something or responsibility for some action. To disavow is not *always* to do something wrong, since in some circumstances one might be called on actively to disavow something, to explicitly condemn and repudiate something with which one has been associated. But in general, I take it that disavowal is a moral failing, insofar as to accuse someone of disavowal is to suggest that they are shirking responsibility, or refusing to own up to or properly respond to something. For instance, we could think here of Donald Trump's repeated disavowal of any role his tweets and speeches might have played in inciting violence, or in encouraging and pandering to his supporters on the extreme right. In such cases, the act of disavowal – the refusal to accept or recognize any responsibility for the (likely) effects of one's words – is itself a moral failure, not (just) an epistemic one.

Butler is primarily concerned with this final sense of denial (viz. disavowal), whereas the sense Manne apparently has in mind (viz. a narrowly epistemic form of denial) is of relatively little interest. Butler emphasizes this point when they write that "there is no thinking of life that is not precarious – except, of course, in fantasy, and in military fantasies in particular" (*FW* 25). They make a similar point in a discussion of vulnerability: "This is a condition, a condition of being laid bare from the start and with which we cannot argue. I mean, we can argue with it, but we are perhaps foolish, if not dangerous, when we do" (*PL* 31). It is striking that both passages begin with a claim

even stronger than Manne's. It is not that nobody *would* deny our condition, but that in a sense nobody *could*: we 'cannot argue', there is 'no thinking' of life that is not precarious. Yet for Butler, it is equally important to emphasize that in another sense, we can and often do put up resistance to such a self-conception, albeit at the cost of a flight into 'foolishness' and fantasy – and not an innocent fantasy, but a 'military' and 'dangerous' one.

The kind of denial operative in fantasy is closer to disavowal than to 'intellectual' denial or a lack of knowledge. In making this contrast clear, it will be useful to consider a distinction drawn by Stanley Cavell between knowledge and acknowledgement. The former, on Cavell's view, is a precondition of the latter:

> From my acknowledging that I am late it follows that I know I'm late (which is what my words say); but from my knowing I am late, it does not follow that I acknowledge that I'm late – otherwise, human relationships would be altogether other than they are. One could say: Acknowledgement goes beyond knowledge. (Goes beyond not, so to speak, in the order of knowledge, but in its requirement that I do something or reveal something on the basis of that knowledge.)[44]

Cavell's formulations bring out the normative dimension of acknowledgement. If I am in a position to acknowledge some fact that I know (that I am late, or indeed vulnerable), a 'requirement' is thereby placed on me to respond in some way to that knowledge – 'to do or reveal something'. Cavell's example also serves to underline the fact that I make my acknowledgement *to* someone – though not necessarily someone else, since I can also acknowledge something to myself. This suggests that to be confronted *by* another with a fact that in some way calls for acknowledgement is to be

[44] Cavell (2015: 237).

beholden to a requirement to respond. The way in which we respond will have more or less far-reaching consequences. Cavell notes that when we say, in sympathy, "I know you are in pain", this is precisely a way of acknowledging the other's suffering *to* them, of responding to the claim their suffering makes on us.[45]

I would suggest that for Butler, our ontological condition demands acknowledgement in just this way, acknowledgement that cannot be achieved if we deny that condition. To disavow it in 'military fantasies' is to fail to acknowledge the ethical claim it places on us. We can take up various lived relations to our knowledge: some of those will amount to acknowledging our shared condition, while others will involve disavowing it, repressing it, or alienating ourselves from our knowledge of it.[46] My suggestion is that for Butler, insisting on our condition is a way of demanding that we respond to it truthfully, without evasion. Although Butler, like Cavell, recognizes that "one may well register and resist the claim", they maintain that if we become genuinely responsive to it, "it will become less easy to accept violence as a taken-for-granted social fact" (*FW* 166-7). Butler's ontology is therefore meant to support ethical and political demands by helping us to register the claim made on us by our interdependency and vulnerability. Butler's point is that a genuine responsiveness to our condition will involve nonviolent modes of behaviour. As they put it, "Nonviolence would, then, be a way of

[45] Cavell (2015: 243).

[46] Gilson also highlights the need for acknowledgement of precariousness on Butler's account (2014: 43), and provides a valuable analysis of the costs and dangers of its disavowal; see Gilson (2011); (2014), Chapter 3. However, I take her account to differ from my own insofar as Gilson equates avowal and acknowledgement of something with "being aware of" it (2014: 61), thus understanding it in a primarily 'intellectual' or epistemic manner. On my account, however, genuine acknowledgement calls for expression in action, in particular lived relations to our knowledge. Furthermore, violent modes of action are not (just) *enabled* by disavowal, but themselves amount to a form of it. In that sense, the Cavellian conception of acknowledgement I attribute to Butler is more demanding than the one Gilson seems to have in mind. This helps to show why genuine acknowledgement will *already* involve an ethical response.

acknowledging that social relation [. . .] and of affirming the normative aspirations that follow from that prior social relatedness" (*FN* 9). Furthermore, it is not just in individual ethical action that we can acknowledge (or disavow) our condition, but also through our political institutions: "The political organization of life itself requires that interdependency – and the equality it implies – is acknowledged through policy, institution, civil society, and government" (*FN* 46-7).[47]

A comparison with Simone de Beauvoir might be helpful in elucidating Butler's view. Beauvoir adopts "the perspective [...] of existentialist morality," according to which subjects (human beings) posit themselves as a transcendence – they exist as "an autonomous freedom".[48] On the existentialist picture, freedom (transcendence) is inalienable. But if all humans are fundamentally free, and if a successful ontology demonstrates this condition to us, how can there be such a thing as a liberatory politics? The answer for Beauvoir is that a subject can take up different relations to their freedom. There are both authentic and inauthentic modes of living one's freedom. From the standpoint of 'existentialist morality', then, specific ontic possibilities, those that involve "lapses into immanence", are "an absolute evil", whether welcomed by the subject (complicity) or imposed from without (oppression).[49] We know that we are free, but there is a great "temptation to flee freedom", to take the

[47] If it seems odd to talk about governments, institutions, &c. being capable of acknowledgement or in a state of disavowal, one need only think of Britain's disavowal of its colonial past, or the US police force's failure to acknowledge its institutionalized racism. Judgements about disavowal and acknowledgement need not always be judgements about the conduct of individual moral agents.

[48] Beauvoir (2010: 17). As Nancy Bauer convincingly argues, the phrase 'existentialist morality' or 'existentialist ethics' (*la morale existentialiste*) signals a departure from Sartre's existentialism (2001: 136-8). Although Bauer does not frame her discussion of the differences between Sartre and Beauvoir's respective positions in terms of the notion of authenticity, I take Bauer's account to be consonant in many respects with the view I sketch in the main text.

[49] Beauvoir (2010: 17).

"easy path" of embracing immanence and making oneself "into a thing", albeit never completely.[50] Equally, there are powerful motivations for denying the freedom of others. Beauvoir's point could thus be reframed in Cavellian vocabulary: *knowing* that we are free is in one sense easy, insofar as all of our possibilities for action testify to our freedom. But properly *acknowledging* freedom – one's own, and that of others – is extremely difficult in the face of anxiety and in the situation of oppression.

Women's emancipation, on Beauvoir's view, can therefore only be achieved through particular modes of acknowledgement: the concrete positing of the subject's transcendence and the overcoming of the external limitations placed on her freedom. Progress is to be evaluated with respect to "the individual's possibilities, defined not in terms of happiness but in terms of freedom".[51] But while Beauvoir takes up an analogue of the notion of authenticity, I would suggest that she understands that notion rather differently than Sartre or Heidegger did, in a way that is deeply connected with her project of liberation. In particular, the relationship between what we might call 'external behaviour' and authentic acknowledgement is construed differently. For Sartre and perhaps – as I suggested in Chapter Two – Heidegger, though 'authenticity' and its opposite ('bad faith'; 'inauthenticity') are surely evaluative terms, the distinction is essentially orthogonal to familiar moral and political evaluative distinctions. It is not the 'concrete content' of the external behaviour that determines whether an existentiell possibility is authentic, but rather the particular way in which that possibility is taken up, lived. The upshot of Sartre's waiter example is presumably *not* that waiting tables is essentially inauthentic, but rather that to take up a particular

[50] Beauvoir (2010: 15).
[51] Beauvoir (2010: 17).

subjective attitude to one's social roles – for instance, to "attempt to realize [...] a being-in-itself of the café waiter", as if one were not always free to transcend this role – is to inhabit them inauthentically, to be in bad faith.[52]

At the end of *Being and Nothingness*, Sartre explicitly defers the project of providing an existentialist ethics. But it is unclear what room is left for him to recognize that the content of particular choices (rather than the subject's relation to them) might make a difference regarding the authenticity of that choice, or for him to view particular political situations as genuinely restrictive of subjects' freedom. As Nancy Bauer argues, "the idea that the choice of 'immanence' over 'transcendence' might be *inflicted* on a person – that oppression can be, as it were, genuinely oppressive – is entirely absent in the early Sartre".[53] Beauvoir, on the other hand, recognizes that the 'lapses into immanence' that function as an analogue to inauthenticity on her view are not essentially disconnected from *specific* political situations and modes of behaviour. Sartre's view abstracts entirely from the fact that "the feminine condition" is characterized by an identification of femininity with immanence.[54] It is the "myth of the Woman, of the Other" and the modes of behaviour founded on that myth that prove determinative for the asymmetry in the relative standing of men and women.[55] "Women's drama", as Beauvoir puts it, "lies in this conflict between the fundamental claim of every subject, which always posits itself as essential, and the demands of a

[52] Sartre (2003: 83).
[53] Bauer (2001: 138).
[54] Beauvoir (2010: 17).
[55] Beauvoir (2010: 14).

situation that constitutes her as inessential".[56] Thus, on her view the question of freedom and authenticity cannot be separated from questions of politics.

Now, it is clear that Butler would contest the terms of appraisal invoked by Beauvoir here: Butler rejects "moves toward authenticity as a way of doing politics" in no uncertain terms (*RVR* 25). The shift from an individualistic conception of freedom to a relational understanding of the subject necessitates a shift in focus from 'the individual's possibilities' to the general shared conditions of liveable life – the social 'ties' that make life possible.[57] What must be acknowledged is not (just) freedom, but our constitutive interdependency. Our "guiding social ideal" is "the equal claim to a liveable and grievable life" (*FN* 24), a claim whose acknowledgement involves striving to oppose the present, highly unequal distribution of 'ontic' violence and precarity. In this sense, while the terms of appraisal are different, the complex relationship between our ontology of the human being and the ethical and political tasks that confront us takes on parallel forms in the work of Beauvoir and Butler. For Butler, as for Beauvoir, genuine acknowledgement of our ontological condition demands a transformation of our political situation.

Finally, allow me to consider one more variation on this theme. Butler writes that we have "a primary vulnerability to others, one that one cannot will away without ceasing to be human" (*PL* xiv). There are two points here. The first is the now familiar ontological claim that vulnerability to others is *constitutive* of what it is to be living,

[56] Beauvoir (2010: 17).
[57] It is nonetheless an open question whether the 'existentialist' conception of the self, properly construed, is as individualistic as Butler seems to suggest. If it is not, the shift will be a matter of emphasis rather than a radical departure.

embodied, and therefore human. We can make no sense of the notion of a human being who is not vulnerable to others. In this sense, willing away one's vulnerability is a literal impossibility. If, *per impossibile*, such an act of will were possible, one would cease to exist as human, so the desire for such an act is always mere fantasy. But there is an ethical point here too, which plays on another use of the term 'human'. 'Human' is contrasted here not just with the *non*-human, but with the *in*human – hence Butler's critical claim that "we have yet to become human" (*PW* 100). In giving in to a fantasy of sovereign independence, one risks abnegating one's humanity in the second sense, becoming inhuman: sovereignty is performed loudly and violently in the domination of the other, and the motivation for this fantasy arises in part from the anxiety that attends our vulnerability to others.[58]

[58] Butler draws attention to the ways in which anxiety about one's vulnerability to others might underlie violent actions. This is evident in Butler's analysis of the political aftermath of 9/11. This "public display of our physical vulnerability" (*PL* 7) prompted recognition "that the national border was more permeable than we thought. Our general response [was] anxiety, rage; a radical desire for security, a shoring-up of the borders against what is perceived as alien" (*PL* 39). This 'shoring-up' took the form of a reassertion of sovereignty through military action, imperilling civilian populations in Iraq, Afghanistan, and elsewhere. Nonetheless, Butler also makes room here for a very different response to the anxiety attending injurability: "one can even experience that abhorrence, mourning, anxiety, and fear, and have all of these emotional dispositions lead to a reflection on how others have suffered arbitrary violence at the hands of the US" (*PL* xiv). This might prompt a politics more responsive to the ethical obligation "to stop [the] further dissemination" of violence (*PL* 8). Another case Butler identifies is a more general anxiety about unstable boundaries, and a consequent terror of 'pollution'. Butler uses this idea to diagnose homophobic responses to the AIDS crisis, and the portrayal of homosexuality itself as "a site of danger and pollution" (*GT* 180; see further 178-83). The body's *openness*, the potential for the resignification of the "fixed sites of permeability and impermeability" that underlie "stable bodily contours" (*GT* 180), becomes a site of anxiety and is violently policed.

VI. Two Failures of Acknowledgement: Individualism and the Politics of Vulnerability

We are now in a position to see that while the relationship between ontology and normativity is not straightforward, on Butler's view the two are not straightforwardly separable either. Their ontological claims are not trivial: though 'self-evident', they are constantly disavowed, and acknowledging them involves maintaining a certain practical relation to them. This means that, even though Butler gives ontology a certain 'priority', they are nonetheless not trying to straightforwardly derive an 'ought' from an 'is'.[59] The normativity Butler is concerned with is located in the relation of acknowledgement: to act in ways that acknowledge our ontological condition is ethically better than acting in ways that disavow it. Furthermore, our mode of acknowledging (or failing to acknowledge) our ontological vulnerability will have various downstream consequences for the ontic possibilities and ideals that are salient and attractive to us. Disavowal is an ethical failing in itself, but it also encourages us to remain in the grip of unattractive and illusory ideals.[60]

---

[59] Cf. Murphy's discussion of the relationship between ontology and ethics (2011: 588-9). Murphy seems to assume that a problematic derivation of an 'ought' from an 'is' is the only model we have for supporting normative claims with ontology. On my account, however, the demand to *acknowledge* one's ontological condition bridges this gap, generating ethical obligations. This helps to answer the critical question Murphy raises regarding "what in Butler's description of precariousness motivates [our] obligations" to others (2011: 583). I am not certain how Murphy's account, on which ontology and ethics are "intertwin[ed]" rather than either having (explanatory?) priority over the other (2011: 589), answers that question satisfactorily.

[60] Erinn Gilson's account of the 'ideal of invulnerability' is highly relevant here: see Gilson (2011) and (2014), Part II. Gilson argues that this ideal is the product of a form of "willful ignorance" (2011: 313). On her account, denial of one's vulnerability is "motivated by the desire – conscious or not – to maintain a certain kind of subjectivity privileged in capitalist socioeconomic systems, namely, that of the prototypical, arrogantly self-sufficient, independent, invulnerable master subject" (2011: 312). Gilson, like Butler, views this desire as ultimately illusory and unachievable, so this self-image must "be continually shored up" through a disavowal of vulnerability if we are to be able to continue to ignore the features of our existence that "might unsettle us" (2011: 313). I am grateful to an anonymous referee at *Ergo* for

However, so far I have said relatively little about what makes a particular lived relation to one's ontological condition good or bad. In this concluding section, I shall try to give a preliminary answer to this question, first by making some general remarks about the status and aspirations of Butler's claims, and then by analysing two particular political stances that Butler finds wanting: individualism, with its attendant fantasy of invulnerability, and a 'politics of vulnerability' that seeks to turn (I shall argue, *ontic*) vulnerability into an identity and a basis for politics. Ultimately – perhaps surprisingly – it will turn out that *both* these stances assume the incompatibility of vulnerability and genuine agency, and are therefore myopic and politically inadequate. Butler's position, on the contrary, advocates a commitment to social transformation on the basis of a *shared* vulnerability and interdependency.

Butler is responding to an illusory ideal of independence, mastery, and impermeability that they take to be endemic and actively harmful. The negative aspect of their project consists in attempting to loosen the grip of this ideal, in part by telling a more *truthful* story about our condition. Butler's account is responsive to very general facts that are obscured by an individualistic conception of the human being that counterposes agency to dependence on and vulnerability to others. However, the truthfulness of Butler's metaphysical account is only part of the story. For even if they have exposed the illusoriness of certain conceptions of independence and mastery, it might be objected that Butler has not given us an overriding reason to abandon those ideals. Given that Butler recognizes the anxiety produced by our exposure to others and the idea of

---

emphasizing the importance of Gilson's work regarding the 'illusory ideals' I discuss in this section.

incursions into the self, they need an account of why it is not in our interests to persist in the illusions they identify. For this reason, Butler aims to offer not just an alternative, more convincing ontology, but a "counter-fantasy" (*FN* 44). The attractiveness of Butler's picture turns not just on the veridicality or intellectual plausibility of their ontology, but on its potential political efficacy and its capacity to offer us a vision of our shared life with others to which we can genuinely aspire. This is why Butler suggests that their social ontology is "to be understood more as a social imaginary than as a metaphysics of the social" (*FN* 16). Butler aims to show us how deeply unappealing the individualistic ideal turns out to be, and to furnish us with a new, more desirable way of conceiving our relations to others on the basis of the ontological notions of vulnerability and interdependency.

Butler associates the narcissistic fantasy of invulnerability and the disavowal of dependency with "the assertion of masculine impermeability" (*FW* 24).[61] There is an affinity here with an idea developed by Adrienne Rich:

> There is much to suggest that the male mind has always been haunted by the force of the idea of *dependence on a woman for life itself*, the son's constant effort to assimilate, compensate for, or deny the fact that he is "of woman born."[62]

If the dominant normative understanding of masculinity connects it with independence and self-sufficiency, the recoil from the 'haunting' memory of an original dependence is likely to result in ultimately violent disavowal of that condition by the

[61] Butler frames the idea of bodily impermeability as an "impossible" and therefore illusory achievement in *Gender Trouble* (*GT* 182), foreshadowing their later invocations of that notion. This suggests an important continuity of interest between Butler's early work on gender and the body and their later 'ethical turn,' though I cannot pursue this line of thought further here.
[62] Rich (1986: 11).

'male mind'. That repression of the fact of existential dependence on the mother is, on Rich's account, part of the cause of misogynistic violence and the denigration of motherhood under patriarchy. Butler makes a structurally similar point about the relation between the illusory ideal of absolute self-sufficiency and the awareness of original dependency:

> As we reflect back on [our infantile] condition [of dependency] as adults, perhaps we are slightly insulted or alarmed, or perhaps we dismiss the thought. Perhaps someone with a strong sense of individual self-sufficiency will indeed be offended by the fact that there was a time when one could not feed oneself or could not stand on one's own. I want to suggest, however, that no one actually stands on one's one; strictly speaking, no one feeds oneself. (*FN* 41)

The idea of dependency might strike one as 'insulting', 'alarming', and inassimilable ('we dismiss the thought'). But the solution to this psychic trouble is not to repress our awareness of our interdependency, but to see that the fantasies of invulnerability, impermeability, and absolute self-sufficiency are not only untenable, but ultimately undesirable. Sartre again makes for an interesting contrast here. In a famous passage, he writes:

> Pure shame is not a feeling of being this or that guilty object but in general of being *an* object; that is, of *recognizing myself* in this degraded, fixed, dependent being which I am for the Other. Shame is the feeling of an *original fall* [because] I need the mediation of the Other to be what I am.[63]

While Sartre in some sense recognizes my dependence on the Other – on his account, it is only through the other that "I gain my objectness", that the objective side of my

[63] Sartre (2003: 312).

existence can be disclosed – this dependence is essentially something lamentable.[64] It is the source of a sense of "*uneasiness*" and perpetual "*danger*".[65] There is little room on the early Sartrean picture for our dependence on others and our vulnerability to them to take on a positive valence – hence, the infamous claim in *Huis Clos* that "hell is other people!"[66] What is of special interest here is that Sartre conceives of the limitation placed on my freedom by the Other as something essentially degrading. Freedom is figured as the individual's absolute sovereignty over their actions, meaning that the incursion of the Other becomes something inherently menacing and limiting. But this conclusion seems deeply unattractive, and threatens to close us of from the positive possibilities that openness to and dependence on the Other may hold out. It is Butler's aim to make such possibilities imaginatively accessible.

One might think that the view sketched here is a product of specifically Sartrean commitments, rather than a general consequence of individualism. However, I would argue that similar fears and fantasies can be discerned, albeit in a more moderate form, in the work of other philosophers who share little with Sartre other than a commitment to an individualist conception of free choice. For instance, explaining the normative commitments of her liberal feminism, Martha Nussbaum writes:

> Liberalism responds sharply to the basic fact [...] that each person is one and not more than one, that each feels pain in his or her own body, that the food given to A does not arrive in the stomach of B. The separateness of persons is a basic fact of human life; in stressing it, liberalism stresses something experientially true, and fundamentally important. [...] The fundamental entity for politics is a

[64] Sartre (2003: 294).
[65] Sartre (2003: 299).
[66] I am indebted here to Nancy Bauer's account of Sartre's early conception of relations with the Other, and its connection to *Huis Clos*. See Bauer (2001), Chapter 4.

> living body that goes from here to there, from birth to death, never fused with any other.[67]

Nussbaum's claim that 'the food given to A does not arrive in the stomach of B' contrasts sharply with Butler's view that 'strictly speaking, no one feeds oneself'. It is of course true that when A eats, their food will end up in A's stomach, not B's. But if B is, for instance, A's child, and A is deprived of the means of subsistence, B will most likely also go hungry. And if B is a foster, growing inside a gravida, A, B is quite literally sustained by the food A consumes.[68] The situation of pregnancy apparently gives the lie to Nussbaum's claim that no living body is ever 'fused with any other'. It is perhaps significant, then, that she shifts over the course of this passage from talk about 'persons' to talk about 'living bodies'. The person is an individual, but the living body is not always obviously so.

Nussbaum's description of her own pregnancy suggests she might demur here. She recalls her impression of a "distinct separateness", writing that "before even her hair got into the world a separate voice could be heard inside, proclaiming *its individuality or even individualism*" and surmising that her daughter would "be quite outraged by the suggestion that her own well being was at any time merged with that of her mother".[69] Although this passage is clearly rhetorical in intent, it is also striking in its refusal to countenance anything recalcitrant to individualism in the situation of pregnancy.[70]

---

[67] Nussbaum (1997: 10).
[68] I borrow the terminology of 'foster' and 'gravida' from Kingma (2019: 611). See especially fns. 7 and 8 for her justification of her terminological decisions.
[69] Nussbaum (1997: 44, fn. 98, emphasis added).
[70] Nussbaum's depiction of her pregnancy seems to fit neatly into the dominant 'containment view' discussed extensively and criticized by Kingma (2019). Although Butler has not written much about pregnancy, I suspect that they would be sympathetic to the 'parthood view' Kingma defends (that is, the idea that the growing foster is literally part of the gravida).

What makes the suggestion – which Nussbaum describes as "overweening" – that one human being's well-being might literally be merged with (or dependent on) that of another so 'outrageous'? It seems that here a prior commitment to individualism provides the cover for a refusal to acknowledge any sense of human interdependency.

Nussbaum is nonetheless right that in some sense the 'separateness of persons' is a basic fact – and not one that Butler would reject. Rather, I want to suggest, Butler seeks to complicate our understanding of the sense in which persons – or perhaps better, embodied beings – are separate. The apparent separateness of bodies serves to mask the ways in which they constantly depend on other bodies, and, to that extent, are inextricably connected with them:

> For if we accept that part of what a body is (and this is for the moment an ontological claim) occurs in its dependency on other bodies – on living processes of which it is a part, on networks of support to which it also contributes – then we are suggesting that it is not altogether right to conceive of individual bodies as completely distinct from one another; and neither would it be right to think of them as fully merged, without distinction. (*FN* 197)

The upshot of this view is that we can no longer take the individual body as the 'fundamental entity for politics', as Nussbaum does. On the contrary, for Butler, equality – whose realisation is after all also a liberal goal – is "a feature of social relations that depends for its articulation on an increasingly *avowed* interdependency – letting go of the body as a 'unit' in order to understand one's boundaries as relational and social predicaments" (*FN* 45). Even if we could easily settle the *metaphysical* question of the boundaries of the individual, we would still have to grant Butler's point that the individual cannot be the sole locus of *political* concern, because to be an

individual body is to be dependent on conditions *beyond* one's bodily boundaries. To contest the violence endemic in our political arrangements, we must rethink the body, contest our individualistic self-conception, and avow – that is, acknowledge – our interdependency.

Liberal individualism, however, is not the only political stance that Butler seeks to question. They also argue that vulnerability cannot "serve as the basis for a politics" (*FN* 186). This claim might seem surprising, given that I have been arguing throughout that Butler wants to make political claims supported by an ontology that counts vulnerability as a constitutive feature of embodied life. However, these two positions are not in fact incompatible. On the contrary, the 'politics of vulnerability' that Butler opposes is in some respects closer to individualism than to Butler's own position. What Butler finds problematic is the "discourse of 'vulnerable groups'" or "vulnerable populations" that creates "a class of persons who identify primarily with vulnerability" (*FN* 186). Butler has three connected worries about this model for political action. The first is that it licenses paternalistic impositions, supposedly in the interests of 'the vulnerable'. The second is that it implicitly relies on a binary opposition between agency and vulnerability; indeed, this is precisely what licenses paternalistic intervention. Vulnerable groups are figured as *lacking* agency, and the goal of political action is ostensibly for them to attain it. Finally, this position treats vulnerability as a property that an individual may or may not possess, but which is then treated as *definitional* of specific groups of individuals (*RVR* 22-5; *FN* 186-92). The 'politics of vulnerability' may appear to be a way of doing precisely what Butler demands – *acknowledging* our vulnerability. But it does so in a radically distorted way, by

mislocating and misconstruing that condition. To use a phrase of Erinn Gilson's, it conceives of it as a "reified negative property of certain types of individuals".[71] To put the point in the terms developed above, the mistake is to try to ground politics on a particular *ontic* form of vulnerability. This vulnerability may well be real, and it may be politically urgent to address it. But it should not be conflated with our *ontological* vulnerability, which is not a property that individuals may or may not have but a generalized, shared condition of being embodied.

This dynamic is at play in certain forms of response to the COVID-19 pandemic, during which the term 'vulnerable' has been ubiquitous. In a recent lecture, Butler describes how, in the aftermath of the first US lockdown, the response of many young people – those "not *vulnerable* – or so they think"[72] – was to return to the 'normality' of gathering in public without maintaining social distance, drinking together, homosocial touching, &c. Here, the discourse of 'vulnerable populations', far from attuning us to a shared, inherently *relational* vulnerability, instead encourages the fantasy that vulnerability is a property that some people have and others do not – as if the 'vulnerability' of 'vulnerable groups' has nothing to do with the behaviour of those who are 'not' vulnerable. This is not to deny that (ontically speaking) some groups *are* more vulnerable than others, for instance at higher risk if they contract the virus. This is both undeniable and ethically important. The point is rather that the pandemic represented an opportunity to recognize a deeper vulnerability and interdependency. It has made (and continues to make) salient the fact that part of what my body is and what it does

[71] Gilson (2011: 311).
[72] "Judith Butler: The Force of Nonviolence," https://www.youtube.com/watch?v=HN5D9rlkRcA

can only be understood in the context of its "communion"[73] with that which is outside it – a communion that is not always visible or acknowledged.

As Butler puts a similar point elsewhere, the "bodily boundary" is not so much an "end" as it is a "threshold of the person, a site of passage and porosity" (*FN* 16). The regulations suggested in the pandemic make this manifest: the sphere of influence of one's body, one's breathing (roughly: two metres!), invites reflection on the unseen effects of one's body on those of others. Reciprocally, the closeness of others, say, the crush of pedestrians on a narrow street, or the cough of a fellow passenger in a train carriage, suddenly becomes threatening. The situation testifies to general facts about what it is to be an embodied human being that have always been true, even if they are not generally acknowledged. The 'individual' body is always already in communion with others, its boundaries always already porous and permeable. It is therefore not enough to think of ethics and politics in terms only of the actions and fortunes of discrete individuals.

I hope it is now clear that the opposition to Butler's theses – primarily psychic, rather than argumentative – is not only real, but indeed "highly consequential" (see *PL* 19-20). It is incumbent on us – especially on those of us who are 'not' vulnerable – to resist the entrenchment of current patterns of ontic vulnerability and precarity that has been the disheartening (but not inevitable) result of the pandemic. It is also incumbent on us to

[73] This term is drawn from Merleau-Ponty (2012: 219). In Butler's lecture discussing the pandemic, they suggest in passing that we re-read Merleau-Ponty in the light of the current crisis. Part of the value of Merleau-Ponty's phenomenology of the body is perhaps that it contests the natural conception of bodies as individual and discrete, with well-defined boundaries, drawing attention to modes of (inter)action that we do not always recognize.

begin to imagine a social order that will properly acknowledge our interdependent ontological condition, and to recognize that that condition is not something to be regretted or willed away in favour of a fantasy of sovereignty or invulnerability. To live is to be exposed to others, and much of what is valuable in life is only made possible by that exposure.

I have argued, albeit briefly, that Butler is right to diagnose a certain kind of denial as lying behind much of what is troubling in contemporary politics: our disavowal of our ontological condition, understood not just epistemically, but *ethically*, as a failure of acknowledgement. However, even if the cursory arguments developed in this section do not convince you, I hope that I have nonetheless made plausible the broader methodological point that Butler's project need not rest on a conflation between the ontological and the ontic – that there is an interesting path from the ontological to the normative. Disavowal – failure of acknowledgement – is an *ethical* failure. For this reason, the fact that some of our actions count as an acknowledgement of our ontological condition, while others count as a disavowal of it, enables ontological description to support substantive normative claims.

Furthermore, I argued in this section that we can make good on Butler's promissory claim that ontology can provide not just a more truthful metaphysics, but also a 'counter-fantasy'. I have suggested that foregrounding our ontological condition might open up ontic possibilities that would otherwise be closed off, obscured by the fantasies and ideals that involve a tacit disavowal of our condition. Our ontological commitments colour our understanding of our relations to others, and a new ontology

has the potential to transform that understanding. Nonetheless, as the comparison with Beauvoir suggests, accepting these general points need not mean subscribing to Butler's particular political views, or indeed to their ontological claims. Because of this, I hope that the broad argument of this chapter will remain of interest even to those who reject the specifics of Butler's account.

Conclusion: A Humanism of the Body?

By way of a conclusion, allow me to return finally to the relationship between humanism and Butler's bodily ontology. I would suggest that Butler's ontological claims might provide a new, more attractive basis for the former. By drawing attention to genuinely ontological features of embodiment, Butler encourages us to acknowledge the vulnerability, dependency, and precariousness of all human beings. However, insofar as this acknowledgement is a potential source of solidarity, it need not be limited to human beings. As I suggested in Section II, Butler views our corporeal condition as a link between human and non-human animals; in this sense, Butler's work points us towards a basis for humanism that also militates against a narrow anthropocentrism. Nonetheless, this need not mean that Butler's account has nothing illuminating to say about distinctively *human* vulnerability, dependency, and precariousness. For their 'two level' account suggests both an ontological level of commonality and an ontic level of difference. If we acknowledge Butler's ontological claims, we may begin to envision an expansion of solidarity, which might come to encompass embodied beings in general. But we may also learn to attend more closely to the diverse ontic modalities of vulnerability and dependence that characterise our human, animal lives.

Chapter Four
Giving an Account of Ourselves: Humanity, Animality, and the Philosophical 'We'

In the previous chapter, I argued that there is a certain tension in Judith Butler's work between their worries regarding the potential violence of a humanistic or anthropocentric approach to philosophy, and their aim of providing an ontology of the body that will illuminate and help us contest specifically human forms of suffering. When Butler asks, "How does one object to human suffering without perpetuating a form of anthropocentrism that has so readily been used for destructive purposes?" (*FW* 76), I take this question to be a genuine one, one that they continuously grapple with in their recent work. Their worries about anthropocentrism, broadly poststructuralist in inspiration, centre around the notions of universality and exclusion.[1] Two such worries are given clear expression by James Williams: "Can anthropocentrism avoid ideas of the ends and destinies of some humans to the exclusion of others? [...] Can anthropocentrism avoid privileging humans above other animals and nature once the ends of man become the destiny of the world?"[2] If we adopt the pejorative use of 'anthropocentrism' that I advocated in the Introduction, then the answer to these questions will – trivially – be 'no'. However, if we pose analogous questions regarding 'humanism', we can view these questions as genuine. The idea of a humanistic approach that can avoid these exclusionary consequences should not be pre-emptively ruled out. Nonetheless, self-consciousness about the history of humanism is likely to shatter any naïve self-confidence in that project. For, to briefly reiterate the argument of

[1] Compare Chapter Three, Section I above.
[2] Williams (2015: 305).

the Introduction (Section II), that history is a history of exclusion along at least two axes: philosophical conceptions of the human have instituted hierarchies among human beings, and have elevated humans above non-human animals.

How are these worries about exclusion connected to the notion of universality, and to philosophy's pretensions to speak in a universal voice? Williams' discussion can again provide a useful starting point:

> The appeal to 'we' is [...] a reminder of the form of republican and independence declarations. [...] It raises three critical points associated with such declarations: on the existence of a referent for 'we', on the legitimacy to declare, and on the legitimacy and form of what the declaration imposes on others in the guise of universality. Who is the 'we' making this declaration? Do they have the right to declare for all? What are they asking of others and should they do so?[3]

Here, the discussion primarily concerns overtly political documents. But the 'three critical points' apply equally to the claims of philosophy, insofar as it seeks to reserve for itself a claim to universal authority. The worry is that the supposedly universal claims made by philosophers, either about or in the name of 'the human' – supposing, in anthropocentric spirit, that this is the referent of the troublesome philosophical 'we' – will turn out to silently pass over differences between human beings, and to illegitimately impose merely parochial ideals of humanity. Call this the 'false universals' concern. In claiming to speak *for* others, for all, philosophy instead speaks over them. This in turn raises a question about philosophy's right to make such universal declarations, the ground of its legitimacy to speak for all: in short, its *authority*.

[3] Williams (2015: 305-6).

These worries certainly animate Butler's work, and partly underlie their reluctance to make universal claims on behalf of 'all human beings'. Nonetheless, as we saw in the previous chapter, Butler develops an ontology that makes extremely general claims about embodied life, attempting to illuminate the constitutive vulnerability and interdependency of bodies. Furthermore, they use these ideas to identify modalities of violence that afflict human beings specifically. Butler's ontology seems to embody a claim to universality, but it is for all that not obviously illegitimate. If we agree that Butler's claims are authoritative, we might therefore wonder what gives them their authority. Does Butler speak on behalf of a general 'we'? If so, what gives them the right to do so?

The answer, I think, lies in the concept of acknowledgement. Butler is not making a universal declaration or imposition, but inviting us to acknowledge aspects of our lives that in one sense none of us would deny. They thereby invite us to view in a particular light the fact that we are nonetheless sometimes resistant – even violently resistant – to those aspects of ourselves and our relations to others. Butler is asking us to consider whether we can really make sense of the idea of living beings that are not vulnerable and dependent. If we find that we cannot, we will come to see our tendency to disavow that vulnerability and dependency as grounded in fantasy. Butler – self-consciously, I think – takes philosophical writing to be a form of interpersonal address. It does not just make impersonal ethical claims, but involves an ethical address – one which demands the acknowledgement of the reader.

In the previous chapter, I suggested that this notion of acknowledgement represents an interesting point of contact between Butler's work and that of Stanley Cavell. In the first part of this chapter, I shall examine Cavell's Wittgensteinian conception of the authority of philosophy and its relationship to acknowledgement, in order to provide a tentative initial response to Williams' worries. Rather than assuming or relying on the existence of a referent for the 'we', on Cavell's view the existence of a 'we' is always in question. The philosophical appeal to the 'we' need not be a declaration or imposition, but is best understood as an invitation to community, an invitation to make shared sense of things and of ourselves. I elucidate this conception of philosophy by explaining and drawing out some implications of Cavell's comparison between Wittgenstein's philosophical claims and Kantian aesthetic judgements. This comparison sheds light on the character of philosophy's claim to universality, and on the fragility of that claim. For on Cavell's conception, the philosopher and their reader have equal claims to authority. As Stephen Mulhall puts it, "the philosopher's claim to community can only be shown to be justified insofar as his interlocutor acknowledges it."[4]

As I have already suggested in my discussion of Butler, to reconceive of the philosophical 'we' along these lines provides an attractive way of understanding how a philosopher's general claims about the nature of the human being might be or become authoritative. However, one might worry that drawing a close connection between the philosophical 'we' and the notion of acknowledgement leaves little room for a 'we' that includes non-human animals, who are not in a position to acknowledge the claims that philosophers make about them or on their behalf.

[4] Mulhall (2015: 153-4).

This concern is somewhat misplaced, insofar as the 'we' that philosophy might hope to institute – a community of speakers – need not and should not be identical to a broader 'we' – a community of moral or ethical concern. Nevertheless, more must be said about the senses in which moral community with other animals might be possible. For that reason, I shift my focus in the remainder of the chapter to a discussion of Wittgenstein's writings on non-human animals, and their relevance to our understanding of the relationships between humans and other animals. While Wittgenstein's comments on non-human animals are sometimes read as highly anthropocentric, I argue that this impression is mistaken. On the contrary, Wittgenstein's work draws our attention to the resemblances and continuities between human and non-human forms of life and behaviour as much as it acknowledges the differences between them. I motivate this line of thought by providing a close reading of a famous aphorism of Wittgenstein's: "If a lion could talk, we wouldn't be able to understand it" (*PPF* §327). I conclude this discussion with some brief remarks about how Wittgenstein's view of behaviour and the body as expressive of mental life can provide resources for coming to see non-human animals as part of the moral sphere. The failure to acknowledge the mental life of other animals is an ethical failure as much as it as an epistemic one.

I. Claims to Universality: On Philosophical and Aesthetic Judgements

Given its prominence in his later philosophical writings, it is unsurprising that Wittgenstein's use of the first person plural – the 'we' – has been a matter of controversy. The interpretative debate has partly revolved around a question that

parallels the first that we saw James Williams raise above: Who is Wittgenstein's 'we'? Is it empirical (some specific 'we' contrasted with others) or transcendental (a 'we' that encompasses any possible being with whom we could make shared sense of things)? Indeed, would it be a mistake to think that there is a single – or even a central – pattern of use of 'we' in Wittgenstein's work, as opposed to a variety of different uses suited to different contexts?[5]

However, the interpretative debate has also concerned Williams' second question, about 'the right to declare for all'. Indeed, Stephen Mulhall has suggested that much of "the hostility and incomprehension with which Wittgenstein's work is so often confronted in philosophical circles crystallizes around [...] the question of how his willingness to speak for himself (to speak personally rather than impersonally) gives him the right to speak for others."[6] Insofar as this latter question is concerned with Wittgenstein's – and more generally, the philosopher's – *right* to speak for others, it is a question of justification: of the philosopher's authority, and the legitimacy of their claims. In this section, I will argue that a reading of Wittgenstein's 'we' developed by Mulhall and Stanley Cavell provides an illuminating account of the philosopher's claim to speak for others – the philosopher's use of the "universal voice", to borrow a phrase from Kant's third *Critique* (5:216).[7] Moreover, Mulhall and Cavell both give a prominent role to the philosopher's willingness to speak personally, as is suggested by the quotation from Mulhall above. The connection between personal judgment and

[5] For the classic statement of the idealist interpretation, see [Bernard] Williams (1981). For discussion, see Moore (2012), Chapter 10, especially §4; for (both exegetical and philosophical) criticism of Williams' position, see Mulhall (2008a).
[6] Mulhall (2008a: 394-5).
[7] See also Cavell (2015: 87).

the 'universal voice' will become clearer in due course. It is worth noting at the outset, however, that on such a reading the philosopher's 'we' is an unavoidably fragile one. Their claim to community will always depend upon the acknowledgement of their interlocutor.

One central feature of Cavell's reading of the Wittgensteinian 'we' is a comparison between Wittgenstein's grammatical remarks – remarks about 'what we say when…' – and aesthetic judgements. Cavell makes an important point about the kind of agreement these judgments claim: "Though the philosopher seems to claim, or depend upon, severer agreement than is carried by the aesthetic analogue, I wish to suggest that it is a claim or dependence of the same kind" (*MWM* 88).[8] He is thus suggesting that an understanding of the claim to universal agreement or assent embodied by *aesthetic* judgments will shed light on the claim to universality embodied by *philosophical* judgments. This will turn out to be related to the essentially personal nature of both kinds of judgment. I shall begin by unpacking some salient features of Kant's account of aesthetic judgement, drawing on Cavell's discussion of Kant to do so. I will then draw attention to the analogues of these features in Wittgenstein's philosophical method, explaining how they illuminate his usage of the first person plural. The result will be a conception of philosophical authority and of its claim to community that go some way towards answering James Williams' worries.

[8] Cavell (2015: 88).

A. The Two 'Peculiarities' of Aesthetic Judgment

The salient affinities between philosophical claims and aesthetic judgments can be brought out by considering what Kant calls the two 'peculiarities' of the latter:

> The judgment of taste determines its object with regard to satisfaction (as beauty) with a claim to the assent of **everyone**, as if it were objective. (5: 281)
>
> The judgment of taste is not determinable by grounds of proof at all, just as if it were merely **subjective**. (5:284)

I take it that Kant's reason for using the word 'peculiarities' here is that there is a puzzling tension between the features of aesthetic judgments that he identifies. His formulations help to bring that tension into relief: since such judgments claim the assent of everyone, but cannot be established by 'grounds of proof' that could compel that assent, we are tempted to regard aesthetic judgements as both 'objective' *and* 'merely subjective'. Kant's use of 'as if', however, suggests that neither of these labels is quite right.

What then, are we to make of such judgments? Picking up on Kant's use of 'as if', Cavell suggests an explanation of the difference between aesthetic and properly 'objective' judgments in terms of the difference between aesthetic properties and 'ordinary' empirical properties of things. But in doing so, he also gives powerful expression to Kant's – and, if we share it, our – conviction that such judgments are nonetheless not merely subjective:

> Only "as if" because [beauty] cannot be an ordinary property of things: its presence or absence cannot be established in the way ordinary properties are; that is, they cannot be established publicly, and we don't know (there aren't

> any) causal conditions, or usable rules, for producing, or altering, or erasing, or increasing this "property." Then why not just say it *isn't* a property of an object? I suppose there would be no reason not to say this, if we could find another way of recording our conviction that it is one, anyway that what we are pointing to is *there*, in the object; [...] while we know not everyone will agree with us when we say it is present, we think they are *missing something* if they don't.[9]

I shall discuss Cavell's expression of 'our conviction' shortly. For now, however, I will focus on the first reason Cavell offers for denying that beauty is an ordinary property of objects, since it corresponds to Kant's claim that aesthetics judgments cannot be established through grounds of proof.[10] Cavell claims that the presence or absence of beauty – and so the truth or falsity of a judgement about an object's beauty – cannot be established 'in the way ordinary properties are'. Cavell's reference to properties that can be 'established publicly' suggests that he has in mind the empirical features of objects that can either be perceived directly, or known on the basis of the testimony of others, or demonstrated more formally by experiment or proof – in other words, the kind of properties that could in principle be established 'scientifically'.

For Kant, however, there can be no "science of the beautiful" (5:304), and no objective principles of aesthetic judgment. Kant writes that such a principle would have to be "a fundamental proposition under the condition of which one could subsume the concept of an object and then by means of an inference conclude that it is beautiful." But this, says Kant, is "absolutely impossible. For I must be sensitive of the pleasure immediately in the representation of [the object], and I cannot be talked into it by means of any

[9] Cavell (2015: 83, fn. 8, emphasis in original).
[10] The second reason – which is that there are no rules for the production of beautiful objects – also bears an affinity to some of Kant's claims. In particular, it recalls Kant's comments on genius, which he says "cannot itself describe or indicate scientifically how it brings its product into being", but which can be understood as "a **talent** for producing that for which no determinate rule can be given" (5:307-8).

proofs" (5:285). There is something highly compelling about the general argument Kant is making here. For there is something almost comic about the idea of offering anything like a formal proof to persuade someone of the beauty of an object. Nor could it be established by a survey; an aesthetic judgment is not a report that "a thing really does please universally", and to make such a judgment one does not need "to grope about [...] among the judgments of others and first inform [oneself] about their satisfaction or dissatisfaction in the same object" (5:282). Kant's point is that an aesthetic judgment, unlike a scientific one, must be made on the basis of one's *own* experience of the object under consideration. The judgment that a painting or a piece of music is beautiful is not the result of an inference, the conclusion of a syllogism; it is a personal response to an aesthetic object.[11]

Summarising this line of thought, Kant writes that "it is required of every judgment that is supposed to prove the taste of the subject that the subject judge for himself" (5:282). This suggests that there is a sense in which it is correct to call such judgments 'subjective': they must be made on the basis of the personal response of a subject. However, it does not imply that we should regard them as *merely* subjective. To understand why, consider Kant's distinction between the 'taste of reflection' and the 'taste of the senses'. Judgments about the beautiful belong to the former kind, and "demand [...] assent universally", even though "as experience teaches" such assent is not generally to be expected; to the latter belong "merely private judgments" concerning one's "pleasure or displeasure in something", which do not claim the assent of others

[11] To grant this point, we need not accept that the aspect of our response to which we must be sensitive is the *pleasure* afforded immediately by a beautiful (or otherwise aesthetically commendable) representation. I take this to be a further claim, in principle separable from the central claim that aesthetic judgment is grounded in personal response.

even when such assent is, as a matter of fact, "quite extensive" (5:214). The difference is thus to be understood in terms of the fact that one form of judgment, but not the other, has 'a claim to the assent of everyone', as Kant puts it in his formulation of the 'first peculiarity' of aesthetic judgment.

The question then arises as to what grounds – or could ground – this claim to assent. As should be clear from the preceding remarks, Kant does not think that such a claim is grounded on any *fact* of universal assent. On the contrary, he suggests that the fact that everyone (else) judges something to be beautiful is not in itself a ground for *me* to assent to that judgement: "Taste makes claim merely to autonomy. To make the judgments of others into the determining ground of one's own would be heteronomy" (5:282). Instead, when we speak in the "universal voice" of aesthetic judgment, we are effectively *laying claim* to assent. In other words, we think that everyone *should* agree, even if they do not.[12] Kant himself had a story to tell about how our expectation of assent is grounded in our shared subjective constitution, the fact that we share the same faculties of imagination and the understanding (see e.g. 5:290n). But we do not need to endorse what is sometimes called Kant's transcendental psychology in order to see a connection between the idea that our aesthetic judgments are not merely subjective, and our conviction that they are based on something shared – or, perhaps better, something we might hope to share. To say that you are *missing something* that I can see is, in part, to hope that you might come to see what I see, come to share in my

[12] I here assume a normative reading of Kant's 'claim to universal assent', and I take Cavell to do so too, but there is some controversy in the literature over whether this is the right interpretation. See Guyer (1997), Chapter 4 for an alternative reading, according to which the claim is understood as "an ideal prediction": the judgment that "*x* is beautiful amounts to the claim [...] that, under ideal conditions [...] everyone who perceives *x* will take pleasure in it" (1997: 129). See also Kalar (2006), Chapter 1 for criticism of Guyer and defence of a normative reading.

judgment. But how can one so much as hope this if no proof can be offered in favour of that judgment?

It is important to realise here that the fact that we do not offer proofs in aesthetics does not entail that we cannot offer *reasons* for our judgment. We can and do advance reasons and arguments in favour of a particular judgement, and, as Cavell stresses, these reasons in some way require (or at least invite) a response. Cavell imagines two disagreements, the first concerning the pleasantness of the taste of canary wine, the second concerning the merits of a musical performance. In the first case, modelled on Kant's 'taste of sense', the disagreement terminates in a mere difference in taste or preference: one liked it, the other did not. Here, we might say that the ground of the judgement really is *merely* subjective – both in the sense of having its ground in the response of a subject, and in the sense that it is not beholden to any further standard. There is nothing to be said in its favour, beyond the bare appeal to its subjective pleasantness. But the second case, Cavell claims, is different. He imagines the exchange as follows:

> A: He plays beautifully doesn't he?
> B1: Yes; too beautifully. Beethoven is not Chopin.
> Or he may answer:
> B2: How can you say that? There was no line, no structure, no idea what the music was about. He's simply an impressive colorist.
> Now, how will A reply? Can he now say: "Well, I liked it"? Of course he *can*; but don't we feel that here that would be a feeble rejoinder, a *retreat* to personal taste? Because B's reasons are obviously relevant to the evaluation of performance, and because they are *arguable*, in ways that anyone who knows about such things will know how to pursue.[13]

[13] Cavell (2015: 85, emphasis in original).

Now, it is worth acknowledging that Cavell's discussion of this imagined exchange already departs somewhat from Kant's account, though I take it to be deeply Kantian in spirit. Kant focuses almost exclusively on judgements of *beauty*, whereas Cavell's example casually subverts the idea that 'beautiful' is the only – or even the primary – affirmative term of aesthetic evaluation or criticism that we have at our disposal; indeed, we should be ready to take seriously the possibility, evinced by B1, that 'beautifully' might in some contexts be a *negative* term of criticism. (Cavell's greater sensitivity to the particulars of our aesthetic discourse should come as no surprise, given the pervasive influence of Wittgenstein and Austin on his work.)

However, note here that in the passage from Kant that I quoted earlier, he also – whether intentionally or not – leaves room for the possibility that there are other ways (besides 'proofs') in which I could be 'talked into' a particular response to the object, a particular judgement of it. The crucial point is that this will be a matter not of someone offering me an objective proof, but of their directing my attention – getting me to see the object in this or that light, encouraging me to adopt to a new way of seeing it.[14] Indeed, this is suggested by the kinds of reasons to which Cavell's B1 and B2 appeal: we are encouraged to compare the music of Beethoven to the music of Chopin, to

[14] Compare here Iris Murdoch's remarks on the critic's role in aiding our aesthetic perception: "The art critic can help us if we are in the presence of the same object and if we know something about his scheme of concepts. Both contexts are relevant to our ability to move towards 'seeing more', towards 'seeing what he sees'"; see Murdoch (2014: 31). Given her emphasis on the presence of a common object, Murdoch seems to share Kant's view that the 'help' a critic can provide is essentially both personal and interpersonal. There is no question of an impersonal proof of the beauty of an object; on the contrary, Murdoch's remarks suggest that some interpersonal understanding of the *critic*, of their 'scheme of concepts' – and so, we might infer, of their lines of interest, their perspective on things – is an important precondition of our ability to learn something from them. Significantly, Murdoch goes on to suggest that the aesthetic case can help us understand progress in moral perception. As we shall see, Wittgenstein himself draws a similar comparison.

consider what kinds of performance their respective pieces might invite or demand; we are encouraged to try to discern 'line', 'structure', 'an idea of what the music is about' in performance – and to recognise that all of these dimensions are lacking. But there is no objective test for any of these aspects of performance; there is a great deal of leeway in the application of the critical concepts B1 and B2 employ. None of the reasons Cavell suggests can rationally compel me to hear the music differently – but they may, for all that, be *there* to be seen.

This suggests that the impossibility of rationally compelling grounds – of a proof – neither forecloses the possibility of debate, nor makes criticism otiose. But the final court of appeal is an individual – that is, personal – judgement. In defending that judgement, one may – if it is really to be a *defence*, perhaps must – appeal to reasons. But the nature of the appeal will be an attempt to bring one's interlocutor to see that object in a particular light, and thereby to get them to respond to it differently. Despite the differences, Cavell's remarks remain broadly indebted to Kant insofar as they suggest an essential difference between judgements that are advanced merely on the basis of subjective taste (the taste of the senses), and properly aesthetic judgements (the taste of reflection). The latter, we might say, are never *merely* subjective, but always embody a claim to *inter*subjective assent – even when such assent is not to be expected. This is precisely what distinguishes them from the judgements based on the 'taste of the senses', where there is and can be no such claim.

This point is nicely expressed in Cavell's discussion when he suggests that if one appeals to (merely) personal taste in the context of an aesthetic disagreement, this

represents a form of *retreat*: a refusal to engage in the debate in the terms most appropriate to it, a refusal to appeal to the kinds of consideration that are acknowledged by 'anyone who knows about such things' to be relevant. When the taste of reflection is effectively assimilated to the taste of the senses the 'subjective' aspect of aesthetic judgement is not so much neglected as radically misconstrued. Such judgments have an ineliminably subjective aspect not because they are based on the subject's taste, construed either as an unarguable brute given or at best as a set of proclivities that, though unresponsive to reasons, are malleable through behaviouristic means (such as acclimatisation). The 'subjectivity' of such judgments instead consists in the fact that one cannot pass the responsibility for one's judgment onto others, even to supposed 'experts' in such matters. I must, as Kant puts it, judge for myself. An aesthetic judgment is essentially a *personal* response to an artwork. But it does not follow from this that the views of others are simply irrelevant to my judgment, or that my judgment can never reach out to others. I am tempted to say, on the contrary, that the more personal the judgment, the more stake we might have in others sharing it, sharing *in* it. The claim to assent is at once a petition for community; the personal is at the same time interpersonal. As we shall see, this will prove to be an important affinity with philosophical judgement.

B. Analogues to Aesthetic Judgment in Wittgenstein's Philosophical Method

As we have seen, in aesthetic judgments the 'universal voice' is not justified by a pre-established fact of universal assent. Instead, it is a way of laying claim to such assent. The claim is an essentially fragile one; it rests on the expectation – or better, the hope – that one's interlocutor shares, or might come to share, a whole array of 'subjective'

sensibilities and sensitivities: "routes of interest and feeling, modes of response, senses of humor and of significance and of fulfillment, of what is outrageous, of what is similar to what else".[15] In fact, this list comes not from Cavell's discussion of aesthetic criticism, but from his description of the vision of language he finds in Wittgenstein's work. However, the fact that his words are so amenable to projection into this new context is no accident. For as I suggested earlier, Cavell sees a deep resonance between the kind of claim made by aesthetic judgments and the kind of claim made by Wittgenstein's philosophical invocation of the 'we'. Both involve an appeal to community, on the basis of a personal judgment.

In fact, as David Egan has pointed out, what Wittgenstein himself had to say about aesthetics is importantly resonant with his own claims about his philosophical method.[16] Moreover, Wittgenstein's remarks also bear close affinities to the conception of aesthetic criticism that I outlined above. Egan draws attention to a passage from G. E. Moore's report of Wittgenstein's lectures, in which Wittgenstein claims that what "Aesthetics tries to do [...] is to give *reasons*." But the reasons it gives are of a special kind: they consist of "further descriptions" and of comparisons. Thus, aesthetics aims "'to draw your attention to a thing', to 'place things side by side'." Nonetheless, there is a limit to the kind of convergence one can expect from aesthetic discussions: discussion will be at "an end" if one has supplied one's interlocutor with "'reasons' of this sort", but "what you see [...] still doesn't appeal" to them. Finally, Moore reports that towards

[15] Cavell (2015: 48).
[16] See Egan (2019), Section 8.5.

the end of his remarks on aesthetics, Wittgenstein "said that the same sort of 'reasons' were given, not only in Ethics, but also in Philosophy."[17]

To a striking degree, Wittgenstein's remarks about the kinds of reasons that are offered in aesthetics recall his statements about his philosophical method. Both the art critic and the philosopher seek to get their reader or interlocutor to *see* things – whether art or our language – in a particular light; they intend or hope to change our "*way of looking at things*" (PI §144). In both cases, this is in part a matter of assembling the right "*objects of comparison* which, through similarities and dissimilarities, are meant to throw light on" the features to which the critic or philosopher seeks to draw attention (*PI* §130). More specifically, Wittgenstein's point about the critic 'placing things side by side' recalls his own emphasis on the "fundamental significance" of the idea of "a surveyable representation [that] produces precisely that kind of understanding which consists in 'seeing connections'" (*PI* §122).

The aim of producing such understanding testifies to Wittgenstein's sense that the reader would be missing something if they did not see the connections that he does. Nonetheless, like Kant and Cavell, Wittgenstein thought that the 'reasons' one offers in aesthetic discussion "cannot compel agreement", as Egan puts it; it is, in the end, up to one's interlocutor whether they accept or reject those reasons and the way of seeing that they commend.[18] The same, he suggests, is the case in philosophy. This point is related to a conception of philosophical authority that can be found in Wittgenstein's work.

[17] All quotations in this paragraph are from [G. E.] Moore (1955: 19). While Egan does not cite Wittgenstein's comment about ethics and philosophy, he does unpack the comparison between aesthetic judgments and philosophical claims at length.
[18] Egan (2019: 214).

Stephen Mulhall draws attention to Wittgenstein's "repeated disavowal of any body of distinctively philosophical expertise", a disavowal which implies a "conception of philosophical authority as finding its ground in the capacities of any competent speaker."[19] I take it that the passages Mulhall has in mind include those where Wittgenstein suggests that philosophical "problems are solved, not by coming up with new discoveries, but by assembling what we have long been familiar with" (*PI* §109); that the philosopher's work "consists in marshalling recollections" (*PI* §127); and that the phenomena that are relevant to the philosopher's task already lie fully "open to view" (*PI* §§92, 126). Wittgenstein takes the philosopher's task not to be to add to a distinctive body of philosophical knowledge, but to get us to see aright what is already in front of our eyes. As Mulhall suggests, it is a consequence of such a conception that there can be no question of expertise in philosophy. The relevant phenomena are not just open to the view of the philosopher, but are equally familiar to their interlocutor, and to any competent speaker. This means that the philosopher's judgments are necessarily contestable, that their claim to universality, to community, is always a fragile one:

> So understood, a philosopher's judgment of what makes sense has no less, but also no more, authority than that of any speaker; and so his interlocutor has just as much authority to contest or revise that judgment. The philosopher's claim to community can only be shown to be justified insofar as his interlocutor acknowledges it. [This] conception of linguistic community is one which demands not the subordination of individual judgment but its full and free expression. What turns out to be ours is both, and can only be, yours and mine. But who exactly 'we' may be is itself at issue in every philosophical exchange.[20]

[19] Mulhall (2015: 153).
[20] Mulhall (2015: 153-4).

There are several parallels here with the case of aesthetic judgment. In invoking the 'we', the philosopher makes a claim to linguistic community; speaking in the universal voice, they lay claim to the assent of their interlocutor. But like in the aesthetic case, this claim is made on the basis of one's individual judgment. It is in speaking for myself that I can (hope to) speak for you. Whether the philosopher does in fact speak for their interlocutor – indeed, for us – cannot be guaranteed in advance, and it is in this sense that the philosopher's 'we' remains fragile. It requires the acknowledgement of the interlocutor for its justification. Whether it receives it will again be a matter of whether the philosopher shares 'routes of interest and feeling' (and all the rest) with their interlocutor.

Indeed, whether we can find ourselves in the philosopher's words depends on whether we "say and want and imagine and feel and suffer together. If we do not, then the philosopher's remarks are irrelevant to us. Of course he doesn't think they are irrelevant, but the implication is that philosophy, like art, is, and should be, powerless to *prove* its relevance".[21] Cavell seems to suggest here that the philosopher's claim to community has its origin in their conviction that their remarks are relevant not just to them, but to us. The philosopher must thus risk making a personal judgment that they know their interlocutor or reader might not (come to) share or acknowledge. This means that, like the artist or critic, the philosopher "always takes authority before she has it; as such, she is always doing something potentially illegitimate. But if she gets it

[21] Cavell (2015: 89).

right – that is, if she makes sense to us – then it turns out that her action was not illegitimate, but in fact authoritative all along."[22]

These remarks return us to the critical questions with which we began. James Williams worried that the philosopher's use of 'we' might assume, illegitimately, the right to declare for others. It thus risks imposing a conception of who 'we' are that falsely universalises from a parochial point of view. No relatively abstract discussion of philosophical method will be able to answer such worries conclusively; the risk Williams identifies is a real one. However, to approach philosophy in the spirit suggested by Wittgenstein and Cavell's work would be disavow the model of philosophical discourse that Williams implicitly assumes. Wittgenstein's remarks are not so much declarations as they are invitations to look at things and ourselves in a new light. The 'we' does not assume a referent that precedes philosophical exchange, but rather seeks to establish one. This conception of philosophy thus has an important dialogical moment. While the philosopher does claim authority, it is open to their interlocutor to rebut that claim. Philosophical authority thus depends on a dialogical exchange for its legitimacy.

Getting our language and its grammar clearly into view was arguably Wittgenstein's central concern in his later work. As a result, the acknowledgement his claims depend upon is the acknowledgement that what makes sense to him also makes sense to us; what he can imagine, we too can imagine. This may seem a relatively specific

[22] Srinivasan (2018: 1415). Srinivasan finds this conception of philosophical authority expressed in the work of Nancy Bauer, who in turn acknowledges its relation to Cavell's work; see Bauer (2015: 20).

philosophical project. However, I take it that the conception of doing philosophy and of making a claim to philosophical authority that I have been elaborating can be extended, and made the basis of a more general approach to philosophical questions. To return to my central concern, I take it that this conception gives us a model for speaking philosophically about the human. When the philosopher makes claims about what human beings are like, what matters to them and why they matter, these claims are not based on a special body of expertise or authority. The philosopher is articulating a vision of what human life is (or could be) like – a vision that it is essentially open to their interlocutor to accept or reject. Indeed, I would suggest that something like this conception of philosophical authority implicitly underlies some of the claims we saw Butler make in the previous chapter. Butler invites the reader to recognise themselves in their expression of their sense that we simply cannot imagine – cannot make sense of the idea of – a body that is invulnerable, or radically independent of other bodies. They invite us to view that idea, insofar as we are tempted by it, as a kind of fantasy, and to consider what consequences that fantasy might have for our relations to others. Butler clearly has a strong sense of the *relevance* of their ontological claims – of the potential transformative power of their vision of the vulnerability and dependency of human (and non-human) life. But equally, they recognise that they cannot compel their reader to recognise that relevance. Whether their claims turn out to be authoritative, then, is in an important sense a question of whether we acknowledge them.

I have suggested that this conception of philosophical authority provides us with an attractive vision of what philosophical or ontological accounts of the human being are

trying to achieve – of the kind of claim to authority they embody. They seek to establish a 'we' through their interlocutor or reader's acknowledgement of their claims. But if their authority is essentially reliant on a dialogical exchange, one might wonder how beings who cannot understand or acknowledge the philosopher's claims – for instance because they lack linguistic capacities altogether – could be included in the 'we'. In other words, one might wonder whether this conception of the philosophical 'we' implicitly encodes a form of anthropocentrism.

Some philosophers have indeed criticised Cavell and Wittgenstein along these lines. Cary Wolfe has expressed concerns that Wittgenstein's work – and Cavell's interpretations of it in particular – tacitly privileges the human being through its emphasis on language and on the sense that we (humans) make of things. According to Wolfe, Cavell's "rendering of the human 'form of life' over against 'the so-called "lower"' forms" represents a "slippage from human to [problematically] human*ist* and the ethical foreclosure that attends it."[23] Moreover, while the community the philosopher may achieve with their reader "strengthens the shared ethical call of those *within* the [language] game," this is "only at the expense of weakening the ethical call in relation to those who speak in other tongues."[24] However, to draw immediate conclusions about the 'ethical call' from claims about the community of speakers is to

[23] Wolfe (2003: 51). The source of the semi-quotation is Cavell (2013: 42). It is perhaps noteworthy that in the full sentence, which Wolfe cites on the previous page, Cavell actually talks about "differences between the human and so-called 'lower' *or 'higher'* forms of life" (emphasis added). This seems to me to cast doubt on the idea, seemingly implied by Wolfe's one-sided emphasis on the word 'lower', that Cavell is setting the human form of life 'over against' the forms of life of other animals in the sense of instituting a hierarchy of ethical concern. But it would be rather more plausible – and charitable – to view Cavell as questioning the legitimacy of such hierarchical claims – hence his use of 'so-called' and the scare quotes he places around the words 'lower' and 'higher', acknowledgement of which is curiously absent from Wolfe's discussion.

[24] Wolfe (2003: 51).

assume, falsely, that only those whom we can speak to or with have an ethical claim on us. In other words, it is to conflate the 'we' of a possible or actual speaking community and the 'we' of the ethical community. Perhaps this is encouraged by the metaphor of the 'call', which is suggestive of a demand posed in language; perhaps the idea is that we cannot or will not 'hear' a call that is not so posed. But the pained cry of an animal or a human infant – a cry that is not, or not yet, language – obviously also 'calls for' an ethical response.

It is trivially true that philosophy, insofar as it is a form of linguistic address, cannot speak directly to or with those who cannot themselves speak or read. In this sense, the claims philosophers make about non-human animals and human beings who cannot speak will not admit of the same kind of interpersonal authorisation or legitimisation that we have been discussing so far. Nonetheless, the philosoper can be more or less attentive to such forms of life, to their similarities to and differences from our own, to the claim they exert upon us. In the remainder of this chapter, I will shift my focus to some of Wittgenstein's remarks about other animals, and try to show why I think his work provides us with a starting point for a greater attentiveness to their forms of life. This will return us, in a circuitous way, to the questions about the 'we' that I have been considering so far.

II. On (Mis)understanding: Lions, Wittgenstein, and Each Other

In this section I hope to illustrate how the suggestions made so far regarding Wittgenstein's use of 'we' might be substantiated and extended. To do so, I will offer a close reading of a famous – and famously opaque – sentence from the *Investigations* that

is sometimes taken to emphasise a radical difference between humans and other animals: "If a lion could talk, we wouldn't be able to understand it [*ihn nicht verstehn*]" (*PPF* §327). On typical readings of this enigmatic sentence, it is claimed that Wittgenstein is giving powerful expression to the idea that a shared 'form of life' is in some sense a precondition for linguistic comprehension. Indeed, it is sometimes suggested that for some behaviour to genuinely count as speech at all, the form of life in which that behaviour has its place must be sufficiently accessible to us. Put briefly, Wittgenstein seems to be claiming that our comprehension of the lion's (putative) speech is radically limited by the inaccessibility of the lion's form of life.

I will explain two variations on this kind of reading. According to the first, what Wittgenstein is asking us to imagine is a lion that is genuinely speaking, but whose speech would nonetheless be radically unintelligible to us; the lion would have a 'conceptual scheme' that is incommensurable with our own. According to the second, Wittgenstein's point is a kind of *reductio* argument: because we could not in principle understand the lion's 'speech', we could not intelligibly count the noises the lion is making as genuine speech after all. I will then explain the assumptions I take to underlie both readings: briefly put, that the 'we' in Wittgenstein's sentence has a fixed and quasi-transcendental referent; that the lion's (putative) *speech* is what we would fail to understand; and that the 'wouldn't be able to' is meant to denote something that we could not do even in principle – a conceptual impossibility. I argue that when we consider the passage in context, bearing in mind the more open-ended interpretative possibilities concerning Wittgenstein's use of 'we' that I discussed in Section I, none of these assumptions are compulsory. I then offer a reading of the passage with a rather

different moral. Wittgenstein is not trying to reveal an absolute limit to our comprehension of the lion, let alone of non-human animals in general. On the contrary, this sentence and the passages that accompany it seek to draw out a distinction between linguistic understanding – understanding of another's speech – and a broader understanding of others, including animal others. There is a way of failing to understand another that is not merely a failure to understand what they are saying. By drawing on recent work by Alice Crary that brings Wittgensteinian considerations in the philosophy of mind to bear on ethics, I shall suggest that the possibility of such forms of misunderstanding returns us to problems that are primarily ethical, rather than theoretical, in character.

First, allow me to sketch some of the extant readings of the lion passage that I alluded to above. One kind of reading takes Wittgenstein's claim at face value, insofar as it does not question the intelligibility of the scenario we are asked to imagine. We are able to imagine a talking lion. But, on this first reading of Wittgenstein's sentence, what we cannot imagine is understanding what the talking lion is saying. Part of the strangeness of the scenario is that, as Cary Wolfe puts it, we are being asked to "to imagine a language we cannot understand, spoken by a being who cannot [in fact] speak."[25] The ostensibly Wittgensteinian explanation for this notion of a being who can talk but whose language 'we' cannot understand appeals to the idea of 'forms of life'. Wittgenstein suggests that the possibility of linguistic agreement – and therefore of mutual understanding – is grounded in "agreement in [...] form of life" (*PI* §241). On Simon Glendinning's reading of the lion passage, Wittgenstein seems to "add to this

[25] Wolfe (2003: 44). The phrase in square brackets is my own clarificatory interpolation.

[conception] the thought that the forms of life that belong to other animals are incommensurable with our own."[26] The lion's life is "just *too different* from a human life" for us to be able to 'agree' in our speech, and so for the lion's speech to ever be intelligible to us. Thus, for Glendinning, Wittgenstein "stresses a difference of *order* in the 'opacity of the other' that a human being can encounter in its respective relations with other human beings and other animals."[27] The lion – and, as Glendinning seems to suggest, non-human animals more generally – are understood as "*radically* other".[28] Even if they could speak, we would not be able to understand them. The upshot is that "human openness presents [a] kind of 'closed horizon'".[29] The possible language-games of other animals would simply not be commensurable with those of human beings.[30]

A second reading takes this notion of the unintelligibility of the lion's 'speech' one step further, and construes Wittgenstein's claim as a *reductio ad absurdum* of the idea that a lion could talk. In their commentary on the *Investigations*, Gordon Baker and Peter Hacker argue that the reason that we could not understand the putative talking lion is that "any 'form of life' accessible to lions, given their natural repertoire of behaviour

---

[26] Glendinning (1998: 70).

[27] Glendinning (1998: 71). In support of this claim, Glendinning emphasizes the fact that Wittgenstein writes that while we sometimes '*do not*' understand other people, we '*could not*' understand the lion.

[28] Glendinning (1998: 72, emphasis added).

[29] Glendinning (1998: 71). Glendinning suggests that Wittgenstein's gesture here is comparable to Heidegger's peculiar privileging of the human being in his analytic of Dasein. See Glendinning (1998: 61-75, especially 70-71); see further Section III of my Introduction, and fn. 3 to Chapter Two above. Nonetheless, Glendinning also draws attention to what he takes to be a separate and opposing strand in Wittgenstein's thought, which emphasizes the continuities between human and non-human animal life and on which "human life can be comfortably acknowledged as being itself a manifestation of animal nature" (1998: 75). On my reading, the lion passage is not in fact in tension with this strand of Wittgenstein's thought.

[30] It should be noted that Wittgenstein has often been read as arguing for a comparable incommensurability between the different language-games of human beings. On such readings, different human forms of life will ground different language-games, between which communication will – or at any rate, may – be impossible.

and their behavioural dispositions, is too far removed from ours for any noises they might emit to count as speech."[31] We would not understand the lion because, contrary to what Wittgenstein initially seems to be asking us to imagine, we could not even count the sounds the lion is making as speech. In support of this passage, Baker and Hacker cite Wittgenstein's claim that "shared human behaviour is the system of reference by means of which we interpret an unknown language" (*PI* §206). The thought is that in the absence of such behaviour, in the absence of that system of reference, there is simply no basis on which we could interpret the lion's apparent 'speech'. To complete the argument, Baker and Hacker appear to be relying on something like the Davidsonian idea that if putative linguistic behaviour is in principle unintelligible to us, if we can make no sense of the idea of interpreting or translating it, then we no longer have grounds for calling it 'language'.[32] On this alternative reading, the idea that we could really imagine a *talking* lion seems to collapse, revealing itself to be a piece of disguised nonsense.

Despite their differences, these two readings both attribute to Wittgenstein a deep pessimism about the prospects of coming to understand other animals. The point is not

[31] Baker and Hacker (2009: 173, fn. 1).

[32] See Davidson (2001). Adrian Moore has suggested that a Davidsonian line of thought can be detected in PI §207; see Moore (2012: 269, fn. 50). In this passage, Wittgenstein asks us to imagine that we come across people who appear to speak in an unknown language, but whose apparent language we find it impossible to learn, "because there is no regular connection between what they say, the sounds they make, and their activities". Wittgenstein concludes the passage by claiming that "the regularity of what we call 'language' is lacking." The apparent moral is that such a lack of regularity undermines our grounds for calling the apparently linguistic behaviour of these people 'language' at all. It is thus no coincidence that Baker and Hacker's discussion of the lion passage refers the reader to their exegesis of this very section. While the first reading of the lion passage discussed in the main text would appear to be in tension with the argument of §207 (since the very unintelligibility of the lion's 'speech' would seem to undermine its claim to be genuine speech at all), Baker and Hacker's reading seems to cast the lion passage as its corollary.

just that, as an empirical fact, non-human animals lack language, and so lack the principal medium through which we come to understand the desires, needs, and points of view of others. The point, on these readings, is that even if other animals were able to speak, there would still be insuperable barriers to our understanding of what they were saying – or, on the second reading, 'saying'. This second, more severe reading seems to suggest that language marks humans off from other animals not just as a matter of fact, but in principle. Wittgenstein's work, on such a reading, seems to shore up the traditional division between human beings and other animals in terms of the '*logos*' – and, we might worry, the attendant elevation of humans above other animals.[33]

One exegetical reason to resist this line of thought, however, is that in many of his other scattered remarks on non-human animals, Wittgenstein seems to emphasise the continuities and resemblances between human and non-human life as much as he does the differences. He draws our attention to cases in which we readily apply some of our psychological and mental concepts to other animals, cases in which we can 'make ourselves understood' to other animals, and even cases in which animals can be said to possess forms of language, albeit of a 'primitive' kind.[34] Moreover, he also draws our attention to non-linguistic forms of understanding: our understanding of a piece of music or a tone of voice, but also our recognition and comprehension of facial expressions and other forms of expressive behaviour.[35] These broader considerations should give us pause, and prompt us to wonder whether an interpretation on which

[33] See Derrida (2008).
[34] See e.g. *PI* §§25, 284, 647; *PPF* 1; *Z* §390; *CV* 77.
[35] See *PI* §§285, 357, 536; see also *PI* §§244-5, 526-535; *PPF* §§149, 354.

Wittgenstein is positing a radical distance between humans and other animals is really the right one.

In fact, the readings of the lion passage that we have considered so far share several assumptions that may be open to question. First of all, both readings assume that what we would not be able to understand in this imaginary scenario is primarily the lion's (putative) speech. This is encouraged by the ambiguity of the English pronoun 'it': when we read 'If a lion could talk, we wouldn't be able to understand *it*', it is perfectly natural to read 'it' as referring to what the lion is saying. But in fact, the German pronoun '*ihn*' arguably refers to the lion, rather than the lion's speech. This is clearer in Anscombe's earlier translation of the same passage, which reads: '…we could not understand *him*'. Of course, German nouns are gendered, and here the masculine pronoun tracks the gender of the masculine noun '*Löwe*', lion; presumably, the rationale for Hacker and Schulte's revision of Anscombe's translation – 'him' to 'it' – is provided by the fact that English nouns (including 'lion') have no gender. Their amendment is thus grammatically justified.[36] Still, it effectively obscures a distinction that we might want to draw between understanding some being's speech and understanding that being themself.[37]

[36] That said, there may be other issues with opting for the neuter or impersonal pronoun ('it') in the case of non-human animals. As Alice Crary points out in a different context, animal protectionists often oppose the use of that pronoun because of its objectifying connotations – its association with things rather than living beings. See Crary (2016: 8) for a brief but illuminating discussion of this point.

[37] A rather different case in which this kind of distinction is operative in Wittgenstein's work is his famous claim in the *Tractatus* that "anyone who understands me eventually recognizes [my propositions] as nonsensical" (*TLP* 6.54). As Adrian Moore puts it, Wittgenstein here "avoids paradox precisely by casting *himself* as the intended object of understanding, rather than the nonsense, which of course cannot be understood." See Moore (2003: 178).

In this particular instance, it might be objected that this is a distinction without a difference. For while we may have some idea of what it means to understand a human being rather than their speech, do we have a clear idea of what it means to 'understand' a lion? Perhaps not. But even if the answer turns out to be 'no', this fact would itself have some bearing on the interpretation of the passage presently under discussion. Furthermore, it is clear from the context of his remark that Wittgenstein has related questions in mind – questions about the nature of 'understanding' of an interpersonal kind (if not – or not yet – of an interspecies kind). Just prior to the lion passage, Wittgenstein makes the following remarks:

> We also say of a person that he is transparent to us. It is, however, important as regards our considerations that one human being can be a complete enigma to another. One learns this when one comes into a strange country with entirely strange traditions; and, what is more, even though one has mastered the country's language. One does not *understand* [*versteht*] the people. (And not because of not knowing what they are saying to themselves.) We can't find our feet with them. [*Wir können uns nicht in sie finden.*] (*PPR* §325)

Clearly, the kind of understanding at stake when we 'find our feet with someone' comes apart from linguistic understanding, our understanding of someone's language. For Wittgenstein is discussing a case in which another human being is a complete enigma to us, where we do not understand them, *even though* we have mastered their language. One of the reasons Wittgenstein is interested in this kind of case relates to his abiding concern with the way a particular picture of 'the inner' – the mental, and what we are tempted to think of as its special privacy and opacity – takes us in. This is evident in the parenthesised sentence from the passage just quoted: when we find another human being opaque, Wittgenstein is keen to stress, it is not always because we

do not know what they are saying to themselves – what is going on 'in their heads'.[38] If our model of opacity is that of a kind of metaphysical inaccessibility, we risk misconstruing the real problems of interpersonal understanding and misunderstanding.

Where the lion passage occurs in the *Last Writings on the Philosophy of Psychology*, the remark that immediately follows it strongly suggests that Wittgenstein understood the imaginary case of the talking lion to be a continuation of this problematic: "Even if someone were to express everything that is 'within him', we wouldn't necessarily understand him" (*LWPP* I §191). Bearing this in mind, we might view Wittgenstein's lion in a rather different light. For part of his point seems to be that insofar as the lion is beyond our understanding, the addition of speech would not necessarily get us any further. Why not? Because the kind of understanding involved in 'finding one's feet with someone' is in no way guaranteed by one's mastery of their language. If the question is whether (or when) we could be said to understand the lion in this sense, then the question of what the lion is saying – and of whether what is said is literally intelligible to us – is perhaps subordinate to the question of what it would mean to 'find our feet' with the lion.

[38] This aspect of Wittgenstein's interest in our failure to understand one another – and his efforts to avoid misconstruing that failure – is more evident when these passages are read in context; see *PPF* §§323-6. It is perhaps noteworthy that immediately before another occurrence of the lion passage (*LWPP* I §190), Wittgenstein puts a cross-reference to a page of his typescript that the editor identifies as corresponding to *RPP* II §§566-569. In these sections, which overlap somewhat with the remarks that precede the lion passage in Part II of the *Investigations*, Wittgenstein is again concerned with thinking through the relationship of the 'inner' and the 'outer' to questions concerning the transparency or opacity of others. In §568, instead of people who remain opaque to us despite our mastery of their language, Wittgenstein asks us to imagine people who "make themselves inaccessible to me by thinking aloud in a language I don't understand. Now I say 'I have no idea what is going on inside them', and there it is — an external fact." In this case, the equation between the inner and the opaque is once again misleading, though now for a different reason: these peoples' thoughts, emotions etc. have been externalised, spoken aloud, but are for all that still just as inaccessible to me.

It would surely be pertinent to ask at this stage what kinds of things we are meant to imagine the lion saying (though it is perhaps a striking fact that no commentator I know of has explicitly done so). Are we being asked to imagine a lion whose naturally expressive vocalisations – growls, roars, yelps of pain – have been magically replaced by English sentences that perform similar functions? For we could surely find English sentences that are put to an analogous – if not quite an identical – use to those vocalisations. A growl is sometimes a warning, a roar is sometimes a means of intimidation, and a yelp of pain needs no clarification at all. In some respects, we understand the roar of a cornered lion perfectly well – and a 'translation' of that roar into English would add little or nothing.[39]

While it would be fanciful to say that lions really do talk, it would not be fanciful to say that lions really produce sounds that are expressive, meaningful (in a 'natural' if not a linguistic sense), and to some extent comprehensible. To the extent that we have a clear idea of what it might mean for a lion to talk, I am not convinced that we would have no grounds for thinking that we could understand what the lion is saying. It therefore seems unlikely that Wittgenstein's point was to produce a straightforward *reductio ad absurdum* of the very idea of a talking lion.[40] Recall that one piece of support for that

[39] Think also of the ease with which a narrator like David Attenborough can give linguistic expression to the dynamics at play in a lion's hunt or a struggle over territory. I would not want to downplay the constant risk of anthropomorphism that attends that genre of narration – in other words, the risk of not taking the otherness of non-human animals seriously enough. Nonetheless, that risk would be incomprehensible *as* a risk – as a possibility of genuine failure on the part of the documentarian of nature – if we did not think there was *something* to get right, *something* that could be legitimately expressed.

[40] Compare here his hesitancy regarding the case of a talking pot in a fairy tale at *PI* §282. It would surely be somewhat surprising if Wittgenstein were more confident about the inconceivability of a talking lion than that of a talking pot!

interpretation was Wittgenstein's claim that "shared human behaviour is the system of reference by means of which we interpret an unknown language." But how exactly should we interpret *that* claim? Is Wittgenstein claiming that shared human behaviour is the only conceivable system of reference for the task of interpretation? Would his claim remain plausible if we often encountered the 'unknown languages' of beings other than humans? For it is surely a contingent truth – if it is a truth – that human beings are the only animals that speak. If we did know of other kinds of creature with whom we could converse, we would perhaps be inclined to drop the qualifier 'human' in 'shared human behaviour'.

Alternatively, the point might be the weaker one that it would be hard to imagine counting something as a language if it could not be correlated in even a minimal respect with any kind of human behaviour. This weaker point would not only be much more plausible, but would also cohere well with Wittgenstein's claims about the relationship between human behaviour and the attribution of our psychological concepts. For instance, he writes that "only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious" (*PI* §281); "only of what behaves like a human being can one say that it *has* pains" (*PI* §283); "we say only of a human being and what is like one that it thinks" (*PI* §360). But a crucial aspect of all of these claims is the elasticity of the concepts that Wittgenstein employs when extending our psychological concepts beyond the case of human beings: *resembles*; *behaves like*; *is like*. In every case, the degree of resemblance, of likeness, is left open; it is not so much determined as *to be determined*. In contrast, Baker and Hacker's brief remarks about the lion passage provide

a decisive – and in my view, overhasty – determination. Should we really be so confident in the judgement that any conceivable form of life for lions "is too far removed from ours for any noises they might emit to count as speech"? What if we were to include, as Wittgenstein might put it, the most primitive forms of speech?[41]

When I claim that this determination is overhasty, I am not necessarily claiming that it is, all things considered, the wrong thing to say. The point is rather that it seems simply to sidestep all of the difficulties that might be involved in such a decision, and all of the questions Wittgenstein's aphorism seems to invite regarding our concept of understanding. Most of all, it seems to sidestep the serious imaginative work that the statement's terseness and strangeness seem to demand. These issues are only exacerbated when commentators take Wittgenstein's lion to be emblematic of non-human animals in general. For to generalize the apparent moral is to fail to heed the specificity of the example – to fail to ask why Wittgenstein choose to imagine a talking *lion*, as opposed to a talking dog or bird or elephant. It would surely be wrong to think that such choices make no difference at all. For consider this simple variation: "If a dog could talk, we wouldn't be able to understand him/her." There will again be a question regarding what kind of 'talk' (in what kind of context, etc.) we are meant to be imagining here. But if Wittgenstein had written *this* sentence, I suspect that most readers – at least, most readers who have spent a significant amount of time with dogs – would be much more hesitant about taking it as a matter of course. I also suspect that there would be less of a temptation to construe the 'we' in this sentence as a quasi-transcendental 'we', or as a 'we' whose referent has already been fixed. Adrian Moore

[41] Cf. *PI* §25.

suggests in passing that the 'we' in the lion passage includes "any beings, actual or imaginable, with whom humans can communicate".[42] On that interpretation of the 'we', the 'dog variation' would surely be false, even if we might think that there is *some* set of human beings (some imaginable, empirical 'we') for whom the sentence is true – perhaps those human beings who 'cannot find their feet' with dogs and other companion animals.

If Wittgenstein is interested in drawing out 'a difference of *order* in the opacity of the other', as we saw Simon Glendinning suggest above, then it is far from obvious that this difference can be straightforwardly correlated with our 'respective relations with other human beings and other animals'. In a passage from *Zettel*, Wittgenstein asks us to "imagine that the people of a tribe were brought up from early youth to give no expression of feeling of any kind. They find it childish, something to be got rid of" (Z §383). After considering various elaborations of this scenario, he imagines the following exchange:

> 'These men would have nothing human about them.' Why? — We could not possibly make ourselves understood to them. Not even as we can to a dog. We could not find our feet with them.
> And yet there surely could be such beings, who in other respects were human. (Z §390)

Let me note in passing that Wittgenstein's interlocutor's use of 'human' is more or less explicitly normative; it suggests that the speaker finds something inhuman in the training or acculturation that these people undergo. While the scare quotes around the first sentence introduce a certain distance, marking it off as a claim in need of further

[42] Moore (2012: 269).

investigation, it is clear that Wittgenstein is sympathetic to it at least to the extent that he tries to articulate its appeal. His sense of inhumanity here – and, if we share it, ours – turns out to be grounded in our inability to make ourselves intelligible to such people. In one respect, this is a reversal of the problem of opacity we encountered earlier: the disturbing thing about this scenario is not so much that these people are an enigma to us, but that *we* are an enigma to *them*. But it is significant that the phrase 'we could not find our feet with them' reoccurs here. Finally, and most relevantly for my immediate purposes, Wittgenstein contrasts our inability to make ourselves understood to these human beings with our ability to make ourselves understood to a dog. While the phrase 'not even…' perhaps suggests that he thinks this ability is in some respects severely limited, it is nonetheless clear from this passage that Wittgenstein did not think that we are in principle radically opaque to other animals, nor they to us. On the contrary: in certain circumstances, we may find it easier to 'find our feet' with other animals than with other human beings.

This suggests, at a minimum, that Wittgenstein's purpose in the lion passage was not to make a general claim about the impossibility of communication between humans and non-human animals. I am not convinced that the passage is intended to have an easily stated moral at all. The point seems rather to be to call into question – to open up for investigation – our notion of 'understanding'. In particular, it seems to invite us to reflect on what it might mean to 'find our feet' with others – human and non-human – and to consider when and why we might find it impossible to do so. These are not just theoretical questions, but ethical ones.

Conclusion: Towards an Ethical Understanding of Other Animals

What might be involved in answering such questions? I want to conclude by giving a brief indication of how Wittgenstein's work might provide us with a promising starting point. While it is true that his remarks on attributions of mental life might appear to centre the human, insofar as they suggest that we can only say of beings relevantly *like* humans that they have pains, sensations, etc. (*PI* §§281, 283, 360), it is worth dwelling on what he takes the relevant kinds of resemblances and likenesses to be. They are, first and foremost, resemblances in expressive behaviour – behaviour that expresses or can express the mental life of a creature. What makes such expressiveness possible is the possession of a living body: "What gives us *so much as the idea* that beings, things, can feel? [...] Only of what behaves like a human being can one say that it *has* pains. For one has to say it *of a body*" (*PI* §283, final emphasis added). In his account of mental ascriptions, Wittgenstein is thus pointing primarily to what is *common* to humans and other animals as the source of our idea that such beings can feel – what is common to adults, infants, cats, beasts, and even to a wriggling fly (*PI* §§244, 284, 647): "If one sees the behaviour of a living being, ones sees its mind [or soul – *Seele*]" (*PI* §357). This suggests that to begin to 'understand' other animals in the sense discussed in the previous section, we must attend carefully to their behaviour.

Alice Crary has argued that a Wittgensteinian conception of mental concepts brings all animals 'inside ethics', for on such an account "these concepts are ethically inflected categories that, in addition to resisting meaningful reduction or translation to neutral terms, are also metaphysically revelatory in the sense of illuminating how things

empirically are."[43] To understand what behaviour is expressive of pain, of curiosity, of intention, we need to have an understanding of what is important in the lives of animals; in this sense, we have to make use of specifically ethical, non-neutral resources. Moreover, Crary suggests that "there can be no question of limiting the imaginative exercise that we [...] face by specifying ahead of time which aspects of human beings' or animals' lives are of interest." [44] The task involves an ongoing work of moral imagination. It will also involve, as Wittgenstein often emphasises, a method of comparison, of placing forms of behaviour side by side and considering lines of continuity and discontinuity. His philosophical method, then, is potentially revelatory of the kinds of significance that animals' behaviour can have – of the whole web of resemblances, affinities, and differences that mark our kinship to and distance from them.

Finally, and perhaps most basically, let us return to the notion of acknowledgement. As in Butler, there is a persistent suggestion in both Wittgenstein and Cavell that to acknowledge the mind of another – and so their interests, susceptibilities, pains, vulnerabilities, their claims to our sympathy or pity – we must acknowledge their embodiment:

> To withhold, or hedge, our concepts of psychological states from a given creature [...] is specifically to withhold the source of my idea that living beings are things that feel; it is to withhold myself, to reject my response to anything as a living being; to blank so much as my idea of anything as *having a body*. [...] Only I could reach to the other's (inner) life. I (have to) *respond* to it, or refuse to respond. It calls upon me; it calls me out. I have to acknowledge it.[45]

[43] Crary (2019: 244); see also (2016), especially Chapter 2.
[44] Crary (2016: 91).
[45] Cavell (1999: 83-4).

Nothing can compel me to respond to or acknowledge the inner life of others. But where I fail to do so, I am withholding not only my psychological concepts, but my very idea of embodied life. I am disavowing, above all, what is common to all of us, human and non-human.

Acknowledgements

I am grateful to the Arts and Humanities Research Council and to Oxford's COVID-19 Scholarship Extensions Fund for funding that made this research possible.

Special thanks to my supervisors, Adrian Moore and Amia Srinivasan, for their insightful comments on many, many earlier drafts of this material, and for all the suggestions and advice they have given me over the past four years. I have greatly enjoyed working with them, and I could not have asked for more attentive, encouraging, and supportive supervisors.

I am also grateful to Nikhil Krishnan, for all his advice and support; to Alice Crary and Stephen Mulhall for their incisive comments at my Transfer and Confirmation of Status assessments, and for helping me to see the scope of my project more clearly; to my college advisor, Alison Hills, for many useful and encouraging conversations; to Joseph Schear and Mark Wrathall for allowing me to present a version of my chapter on Butler at the Post-Kantian European Philosophy Seminar at Oxford, and for everything I learnt at their graduate classes on Heidegger, Sartre, and Merleau-Ponty; to two anonymous referees at *Ergo* for useful comments on an article based on the Butler chapter; to the 'Contrybutyons Fan Club', Joe Hamilton, Jae Hetterley, and Fergus Horan, with whom I struggled through Heidegger's *Contributions to Philosophy* for the first time in the early months of the pandemic; to Niamh Sauter-Cooke for last-minute German advice; and to the many colleagues and friends at Oxford and beyond who have greatly shaped my ideas, especially to Patricia Cipollitti, Nick Clanchy, Maeve Cooke, Rachel Fraser, Sasha Lawson-Frost, Anil Gomes, Elisabeth Huh, Maya Krishnan, Chiara Martini, Otto Räsänen, Livia von Samson, Bekzhan Sarsenbay, Leo Schneider, Gabriel de Sousa, Zoe Walker, Rob Watt, and George Webster.

Thanks also to my family – Sadie Wearing, Richard Partington, Hailey Wearing, and Alice Wearing – especially for all the love and support you gave me during the pandemic. Special thanks to Sadie Wearing for many illuminating discussions and for all the academic advice she has given me over the years.

Finally, my heartfelt thanks to Sophie Russell for all her encouragement and insight. ✓

Bibliography

Works by Immanuel Kant

Kants gesammelte Schriften (Berlin: Walter de Gruyter, 1900–).

Volumes from the Cambridge Edition of the Works of Immanuel Kant *(1992–):*

Practical Philosophy, ed. and trans. Mary J. Gregor (Cambridge: Cambridge University Press, 1996).

Lectures on Ethics, eds. Peter Heath and J. B. Schneewind, trans. Peter Heath. (Cambridge: Cambridge University Press, 1997).

Critique of Pure Reason, eds. and trans. Paul Guyer and Allen W. Wood (Cambridge: Cambridge University Press, 1998).

Critique of the Power of Judgment, ed. Paul Guyer, trans. Paul Guyer and Eric Matthews (Cambridge: Cambridge University Press, 2000).

Anthropology, History, and Education, eds. Günter Zöller and Robert B. Louden, trans. Mary Gregor, Paul Guyer, Robert B. Louden, Holly Wilson, Allen W. Wood, Günter Zöller, Arnulf Zweig (Cambridge: Cambridge University Press, 2007).

Works by Martin Heidegger

Gesamtausgabe edition (Frankfurt: Klostermann, 1975–).

English translations of volumes from Heidegger's Gesamtausgabe:

Ga5 *Off the Beaten Track*, trans. Julian Young and Kenneth Haynes (Cambridge: Cambridge University Press, 2002).

Ga9 *Pathmarks*, ed. and trans. William McNeill (Cambridge: Cambridge University Press, 1998).

Ga29/30 *The Fundamental Concepts of Metaphysics*, trans. William McNeill (Bloomington: Indiana University Press, 1995).

Ga38 *Logic as the Question Concerning the Essence of Language*, trans. Wanda Torres Gregory and Yvonne Unna (Albany: SUNY Press, 2009).

Ga40 *Introduction to Metaphysics*, revised 2nd ed., eds. and trans. Gregory Fried and Richard Polt (New Haven: Yale University Press, 2014).

Ga49 *The Metaphysics of German Idealism*, trans. Ian Alexander Moore and Rodrigo Therezo (Cambridge: Polity Press, 2021).

Ga65 *Contributions to Philosophy (of the Event)*, trans. Richard Rojcewicz and Daniela Vallega-Neu (Bloomington: Indiana University Press, 2012).

Ga69 *History of Beyng*, trans. William McNeill and Jeffrey Powell (Bloomington: Indiana University Press, 2015).

Ga77 *Country Path Conversations*, trans. Bret W. Davis (Bloomington: Indiana University Press, 2010).

Ga79 *Bremen and Freiburg Lectures: Insight Into That Which Is and Basic Principles of Thinking*, trans. Andrew J. Mitchell (Bloomington: Indiana University Press, 2012).

Other works by Heidegger:

BT *Being and Time*, trans. John Macquarrie and Edward Robinson (Oxford: Blackwell, 1962).

QCT 'The Question Concerning Technology', in *The Question Concerning Technology and Other Essays*, trans. William Lovitt (New York: Harper & Row, 1977), pp. 3-35.

OM 'Overcoming Metaphysics', in Richard Wolin, ed., *The Heidegger Controversy* (Cambridge: MIT Press, 1993a), pp. 67-90.

OAG '"ONLY A GOD CAN SAVE US": *Der Spiegel's* Interview with Martin Heidegger', in Richard Wolin, ed., *The Heidegger Controversy* (Cambridge: MIT Press, 1993b), pp. 91-116.

LH 'Letter on "Humanism"', in William McNeill, ed., *Pathmarks* (Cambridge: Cambridge University Press, 1998), pp. 239-276.

Works by Judith Butler

Abbreviations of selected works by Butler:

GT *Gender Trouble: Feminism and the Subversion of Identity*, Routledge Classics edition, including 'Preface (1999)' (London: Routledge, 2006).

PL *Precarious Life: The Powers of Mourning and Violence* (London: Verso, 2004a).

FW *Frames of War: When is Life Grievable?*, *Radical Thinkers* edition, including 'Introduction to the Paperback' (London: Verso, 2016a).

RVR 'Rethinking Vulnerability and Resistance' in Judith Butler, Zeynep Gambetti, and Leticia Sabsay, eds., *Vulnerability in Resistance*, (Durham: Duke University Press, 2016b), pp. 12–27.

FN *The Force of Nonviolence: An Ethico-Political Bind* (London: Verso, 2020).

Other works by Butler:

'Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory', *Theatre Journal*, Vol. 40, No. 4 (1988): 519–31.

'Endangered/Endangering: Schematic Racism and White Paranoia', in Robert Gooding-Williams, ed., *Reading Rodney King/Reading Urban Uprising* Routledge, 1993), pp. 15-22.

'Restaging the Universal', in Judith Butler, Ernesto Laclau, and Slavoj Zizek, *Contingency, Hegemony, Universality: Contemporary Dialogues on the Left* (London: Verso, 2000), pp. 11-43.

Undoing Gender (London: Routledge, 2004b).

Giving an Account of Oneself (New York: Fordham University Press, 2005).

'A Reply from Judith Butler', *Philosophy and Phenomenological Research*, Vol. 96, No. 1 (2018): 243–49.

'The Inorganic Body in the Early Marx: A Limit-concept of Anthropocentrism', *Radical Philosophy*, Issue 2.06 (2019): 3-17.

Works by Ludwig Wittgenstein

Abbreviations of works by Wittgenstein:

TLP *Tractatus Logico-Philosophicus*, trans. D. F. Pears and B. F. McGuinness (London: Routledge, 1974).

RPP II *Remarks on the Philosophy of Psychology*, Vol. II, eds. G. H. von Wright and Heikki Nyman, trans. C. G. Luckhardt and Maximilian A. E. Aue (Oxford: Basil Blackwell, 1980).

Z *Zettel*, 2nd ed., eds. G. E. M. Anscombe and G. H. von Wright, trans. G. E. M. Anscombe (Oxford: Basil Blackwell, 1981).

LWPP I *Last Writings on the Philosophy of Psychology*, Vol. I, eds. G. H. von Wright and Heikki Nyman, trans. C. G. Luckhardt and Maximilian A. E. Aue (Oxford: Basil Blackwell, 1982).

CV — *Culture and Value*, 2nd ed., eds. G. H. von Wright, Heikki Nyman, and Alois Pichler, trans. Peter Winch (Oxford: Basil Blackwell, 1998).

PI — *Philosophical Investigations*, 4th ed., eds. P. M. S. Hacker and Joachim Schulte, trans. G. E. M. Anscombe, P. M. S. Hacker, and Joachim Schulte (Oxford: Wiley-Blackwell, 2009).

PPF — 'Philosophy of Psychology – A Fragment', in P. M. S. Hacker and Joachim Schulte, eds., *Philosophical Investigations* (Oxford: Wiley-Blackwell, 2009), pp. 182–243.

Other Cited Works

Theodor W. Adorno, *Problems of Moral Philosophy*, ed. Thomas Schröder, trans. Rodney Livingstone (Cambridge: Polity Press, 2000).

Lucy Allais, 'Kant's Racism', *Philosophical Papers* Vol. 45, Nos. 1 and 2 (March and July 2016): 1-36.

Henry Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990).

Louis Althusser, *For Marx*, trans. Ben Brewster (London: Verso, 2005).

G. E. M. Anscombe, 'Modern Moral Philosophy', *Philosophy* Vol. 33, No. 124 (1958): 1-19.

G. E. M. Anscombe, *Human Life, Action and Ethics: Essays*, eds. Mary Geach and Luke Gormally (Exeter: Imprint Academic, 2005).

G. P. Baker & P.M.S. Hacker, *Wittgenstein: Rules, Grammar, and Necessity – Essays and Exegesis of §§185-242*, 2nd extensively rev. ed. by P.M.S. Hacker (Chichester: Wiley-Blackwell, 2009)

Thomas Baldwin, 'The Humanism Debate' in Brian Leiter and Michael Rosen, eds., *The Oxford Handbook of Continental Philosophy* (Oxford: Oxford University Press, 2007), pp. 671-710.

Nancy Bauer, *Simone de Beauvoir, Philosophy, & Feminism* (New York: Columbia University Press, 2001).

Nancy Bauer, *How to Do Things with Pornography* (Cambridge: Harvard University Press, 2015).

Simone de Beauvoir, *The Second Sex*, trans. Constance Borde and Sheila Malovany-Chevallier (London: Vintage Books, 2010).

Miguel de Beistegui, 'The Transformation of the Sense of Dasein in Heidegger's *Beiträge zur Philosophie (Vom Ereignis)*', *Research in Phenomenology*, Vol. 33, No. 1 (2003): 221-246.

Seyla Benhabib, 'Ethics without Normativity and Politics without Historicity: On Judith Butler's *Parting Ways: Jewishness and the Critique of Zionism*', *Constellations*, Vol. 20, No. 1, (2013): 150–63.

Rosi Braidotti, *The Posthuman* (Cambridge: Polity Press, 2013).

Alexis Burgess and David Plunkett, 'Conceptual Ethics [I and II]', *Philosophy Compass*, 8(12) (2013): 1091-1110.

Stanley Cavell, *The Claim of Reason: Wittgenstein, Skepticism, Morality, and Tragedy* (Oxford: Oxford University Press, 1999).

Stanley Cavell, *This New Yet Unapproachable America: Lectures after Emerson after Wittgenstein* (Chicago: University of Chicago Press, 2013).

Stanley Cavell, *Must We Mean What We Say?*, Updated ed. (Cambridge: Cambridge University Press, 2015).

Robin Celikates, 'Forms of Life, Progress, and Social Struggle', in Amy Allen and Eduardo Mendieta, eds., *From Alienation to Forms of Life: The Critical Theory of Rahel Jaeggi* (University Park, PA: Penn State University Press, 2018), pp. 137-55.

J. M. Coetzee et al., *The Lives of Animals*, ed. Amy Gutmann (Princeton: Princeton University Press, 1999).

Alice Crary, *Inside Ethics: On the Demands of Moral Thought* (Harvard: Harvard University Press, 2016).

Alice Crary, 'Comments on a Contested Comparison: Race and Animals', in Benjamin De Mesel and Oskari Kuusela, eds., *Ethics in the Wake of Wittgenstein* (New York: Routledge, 2019a), pp. 233-55.

Alice Crary, 'Animals, Cognitive Disability and Getting the World in Focus in Ethics and Social Thought: A Reply to Eva Feder Kittay and Peter Singer', *Zeitschrift für Ethik und Moralphilosophie* Vol. 2, No. 1 (2019b): 139-46.

Elizabeth Cykowski, *Heidegger's Metaphysical Abyss: Between the Human and the Animal* (Oxford: Oxford University Press, 2021).

Bret W. Davis, *Heidegger and the Will: On the Way to Gelassenheit* (Evanston, Ill.: Northwestern University Press, 2007).

Jacques Derrida, 'The Ends of Man', *Philosophy and Phenomenological Research*, Vol. 30, No. 1 (1969): 31-57.

Jacques Derrida, *Of Spirit: Heidegger and the Question*, trans. Geoffrey Bennington and Rachel Bowlby (Chicago: University of Chicago Press, 1989).

Jacques Derrida, *The Animal That Therefore I Am*, trans. Marie-Louise Mallet (New York: Fordham University Press, 2008).

Cora Diamond, 'Losing Your Concepts', *Ethics* Vol. 98, No. 2 (1988): 255-277.

Cora Diamond, 'Bernard Williams on the Human Prejudice', *Philosophical Investigations* Vol. 41, No. 4 (2018): 379-398.

Donald Davidson, 'On the Very Idea of a Conceptual Scheme', reprinted in his *Inquiries into Truth and Interpretation*, 2nd ed. (Oxford: Clarendon Press, 2001), pp. 183-198.

Sue Donaldson and Will Kymlicka, *Zoopolis: A Political Theory of Animal Rights* (Oxford: Oxford University Press, 2011).

David Egan, *The Pursuit of an Authentic Philosophy: Wittgenstein, Heidegger, and the Everyday* (Oxford: Oxford University Press, 2019).

Rainer Forst, 'Noumenal Alienation: Rousseau, Kant and Marx on the Dialectics of Self-Determination', *Kantian Review*, Vol. 22, No. 4 (2017): 523-51.

Michel Foucault, *The Order of Things* (London: Routledge, 2002).

Nancy Fraser, 'Climates of Capital', *New Left Review*, 127 (2021): 94–127.

Fabian Freyenhagen, *Adorno's Practical Philosophy: Living Less Wrongly* (Cambridge, Cambridge University Press: 2013).

Patrick Frierson, *Kant's Empirical Psychology* (Cambridge: Cambridge University Press, 2014).

Erich Fromm and Karl Marx, *Marx's Concept of Man*, trans. T. B. Bottomore (London: Bloomsbury Academic, 2013).

Norman Geras, *Marx and Human Nature: Refutation of a Legend* (London: Verso, 2016).

Erinn Gilson, 'Vulnerability, Ignorance, and Oppression', *Hypatia*, Vol. 26, No. 2 (2011): 308–32.

Erinn Gilson, *The Ethics of Vulnerability: A Feminist Analysis of Social Life and Practice* (London: Routledge, 2014).

Simon Glendinning, *On Being with Others: Heidegger, Wittgenstein, Derrida* (London: Routledge, 1998).

Sacha Golob, 'Heidegger', in Sacha Golob and Jens Timmermann, eds., *The Cambridge History of Moral Philosophy* (Cambridge: Cambridge University Press, 2017), pp. 623-635.
Sacha Golob, 'Is Heidegger's History of Being a Genealogy?', *The Monist*, Vol. 105, No. 4 (2022): 507–520.

Paul Guyer, *Kant and the Claims of Taste*, 2nd ed. (Cambridge: Cambridge University Press, 1997).

G. W. F. Hegel, *Lectures on the Philosophy of World History: Introduction*, trans. H. B. Nisbet (Cambridge: Cambridge University Press, 1975).

G. W. F. Hegel, *The Phenomenology of Spirit*, trans. Terry Pinkard (Cambridge: Cambridge University Press, 2018)

Rahel Jaeggi, *Critique of Forms of Life*, trans. Ciaran Cronin (Cambridge, MA: The Belknap Press of Harvard University Press, 2018).

Harriet McBryde Johnson, 'Unspeakable Conversations, or, How I Spent One Day as a Token Cripple', in *Peter Singer Under Fire: The Moral Iconoclast Faces His Critics*, ed. Jeffrey A. Schaler (Chicago and La Salle, Illinois: Open Court, 2009), pp. 195-205.

Shelly Kagan, 'What's Wrong with Speciesism?', *Journal of Applied Philosophy*, Vol. 33, No. 1 (2016): 1-21.

Brent Kalar, *The Demands of Taste in Kant's Aesthetics* (London: Continuum, 2006).

Aret Karademir, 'Butler and Heidegger: On the Relation between Freedom and Marginalization', *Hypatia*, Vol. 29, No. 4 (2014): 824–39.

Eva Feder Kittay, 'The Personal Is Philosophical Is Political: A Philosopher and Mother of a Cognitively Disabled Person Sends Notes From the Battlefield', *Metaphilosophy* Vol. 40, No. 3/4 (2009): 606-627.

Elselijn Kingma, 'Were You a Part of Your Mother?', *Mind*, Vol. 128, No. 511 (2019): 609–46.

Pauline Kleingeld, 'On Dealing with Kant's Sexism and Racism', *SGIR Review*, Vol. 2, No. 2 (2019): 3-22.

Christine M. Korsgaard, 'Kant's Formula of Humanity', in her *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), pp. 106-132.

Christine M. Korsgaard, *Fellow Creatures: Our Obligations to the Other Animals* (Oxford: Oxford University Press, 2018).

Julia Kristeva, 'Powers of Horror: Approaching Abjection', in *The Portable Kristeva*, ed. Oliver Kelly (New York: Columbia University Press, 2002a), pp. 229-47.

Julia Kristeva, 'From Filth to Defilement', in *The Portable Kristeva*, ed. Oliver Kelly (New York: Columbia University Press, 2002b), pp. 248-63.

R. D. Laing, 'The Obvious', in David Cooper, ed., *The Dialectics of Liberation* (London: Verso Books, 2015), pp. 13-33.

Rae Langton, 'Duty and Desolation', in her *Sexual Solipsism: Philosophical Essays on Pornography and Objectification* (Oxford: Oxford University Press, 2009a), pp. 197-222.

Rae Langton, 'Autonomy-Denial in Objectification', in her *Sexual Solipsism* (Oxford: Oxford University Press, 2009b), pp. 223-240.

Primo Levi, *"If This Is a Man" and "The Truce"* (Harmondsworth: Penguin Books, 1979).

Emmanuel Levinas, *Difficult Freedom: Essays on Judaism*, trans. Seán Hand (London: Athlone Press, 1990).

Moya Lloyd, 'The Ethics and Politics of Vulnerable Bodies', in Moya Lloyd, ed., *Butler and Ethics* (Edinburgh: Edinburgh University Press, 2015), pp. 167–92.

Kate Manne, 'Humanism: A Critique', *Social Theory and Practice*, Vol. 42, No. 2 (2016): 389-415.

Kate Manne, 'Melancholy Whiteness (or, Shame-Faced in Shadows)', *Philosophy and Phenomenological Research*, Vol. 96, No. 1 (2018): 233–42.

Wayne Martin, 'The Semantics of "Dasein" and the Modality of Being and Time', in Mark Wrathall, ed., *The Cambridge Companion to Heidegger and* Being and Time (Cambridge: Cambridge University Press, 2013), pp. 100-128.

Maurice Merleau-Ponty, *The Phenomenology of Perception*, trans. Donald A. Landes (London: Routledge, 2012).

Catherine Mills, 'Undoing Ethics: Butler on Precarity, Opacity and Responsibility', in Moya Lloyd, ed., *Butler and Ethics* (Edinburgh: Edinburgh University Press, 2015), pp. 41–64.

Charles Mills, *Black Rights/White Wrongs: The Critique of Racial Liberalism* (Oxford: Oxford University Press, 2017).

A. W. Moore, *Noble in Reason, Infinite in Faculty: Themes and Variations in Kant's Moral and Religious Philosophy* (London: Routledge, 2003a).

A. W. Moore, 'Ineffability and Nonsense', *Proceedings of the Aristotelian Society, Supplementary Volumes* Vol. 77 (2003b): 169-193.

A. W. Moore, *The Evolution of Modern Metaphysics: Making Sense of Things* (Cambridge: Cambridge University Press, 2012).

A. W. Moore, 'Sense-making From a Human Point of View', in Giuseppina D'Oro and Søren Overgaard, eds., *The Cambridge Companion to Philosophical Methodology* (Cambridge: Cambridge University Press, 2017), pp. 44-55.

G. E. Moore, 'Wittgenstein's Lectures in 1930-33', *Mind*, Vol. 64, No. 253 (1955): 1-27.

Stephen Mulhall, *Inheritance and Originality: Wittgenstein, Heidegger, Kierkegaard* (Oxford: Oxford University Press, 2001).

Stephen Mulhall, '"Hopelessly Strange": Bernard Williams' Portrait of Wittgenstein as a Transcendental Idealist', *European Journal of Philosophy*, Vol. 17, No. 3 (2008a): 386-404.

Stephen Mulhall, *The Wounded Animal: J. M. Coetzee and the Difficulty of Reality in Literature and Philosophy* (Princeton: Princeton University Press, 2008b).

Stephen Mulhall, 'Adrian Moore's Wittgenstein', *Philosophical Topics*, Vol. 43, No. 1/2 (2015): 149-160.

Stephen Mulhall, *The Ascetic Ideal: Genealogies of Life-denial in Religion, Morality, Art, Science, and Philosophy* (Oxford: Oxford University Press, 2021).

Iris Murdoch, *The Sovereignty of Good* (London: Routledge, 2014).

Ann V. Murphy, 'Corporeal Vulnerability and the New Humanism' *Hypatia*, Vol. 26, No. 3 (2011): 575–90.

Martha C. Nussbaum, 'Objectification', *Philosophy & Public Affairs*, Vol. 24, No. 4 (1995): 249-91.

Martha Nussbaum, *The Feminist Critique of Liberalism* (Lawrence, Kansas: University of Kansas, 1997).

Anne Phillips, *The Politics of the Human* (Cambridge: Cambridge University Press, 2015).

Steven Pinker, *Enlightenment Now: The Case for Reason, Science, Humanism and Progress* (London: Penguin Books, 2019).

Richard Polt, 'A Running Leap into the There: Heidegger's "Running Notes on *Being and Time*"', *Graduate Faculty Philosophy Journal*, Vol. 41, No. 1 (2020): 55-71.

Willard V. O. Quine, *From a Logical Point of View: 9 Logico-Philosophical Essays*, 2nd ed. (Harvard: Harvard University Press, 1980).

Adrienne Rich, *Of Woman Born: Motherhood as Experience and Institution*, 2nd ed. (New York: W. W. Norton & Company, 1986).

Jean-Paul Sartre, *Being and Nothingness: An Essay on Phenomenological Ontology*, trans. Hazel Barnes (London: Routledge, 2003).

Jean-Paul Sartre, *Existentialism is a Humanism*, ed. John Kukla and trans. Carol Macomber (New Haven: Yale University Press, 2007).

Thomas Sheehan, 'A Paradigm Shift in Heidegger Research' *Continental Philosophy Review*, Vol. 34, No. 2 (2001): 183–202.

Peter Singer, 'Speciesism and Moral Status', *Metaphilosophy* Vol. 40, No. 3/4 (2009a): 567-581.

Peter Singer, 'An Intellectual Autobiography', in *Peter Singer Under Fire*, ed. Jeffrey A. Schaler (Chicago and La Salle, Illinois: Open Court, 2009b), pp. 1-74.

Peter Singer, 'Reply by Peter Singer', in *Peter Singer Under Fire*, ed. Jeffrey A. Schaler (Chicago and La Salle, Illinois: Open Court, 2009c), pp. 97-101.

Peter Singer, *Practical Ethics*, 3rd edition (Cambridge: Cambridge University Press, 2011).

Nicholas Southwood, 'The Feasibility Issue', *Philosophical Compass*, Vol. 13, No. 8 (2018), e12509.

Amia Srinivasan, 'The Ineffable and the Ethical', *Philosophy and Phenomenological Research*, Vol. 96, No. 1 (2018): 215–23.

Amia Srinivasan, 'What Does Fluffy Think?', *London Review of Books*, Vol. 43, No. 19, 7 October 2021.

P. F. Strawson, 'Freedom and Resentment', in his *Freedom and Resentment and Other Essays* (London: Routledge, 2008), pp. 1-28.

Iain Thomson, *Heidegger, Art, and Postmodernity* (Cambridge: Cambridge University Press, 2011).

Jeremy Waldron, *One Another's Equals: The Basis of Human Equality* (Cambridge: Harvard University Press, 2017).

Ralph C. S. Walker, 'Self and Selves', in Anil Gomes and Andrew Stephenson, eds., *Kant and the Philosophy of Mind: Perception, Reason, and the Self* (Oxford: Oxford University Press, 2017), pp. 204-220.

Bernard Williams, 'Wittgenstein and Idealism', in his *Moral Luck: Philosophical Papers, 1973-1980* (Cambridge: Cambridge University Press, 1981), pp. 144-163.

Bernard Williams, *Philosophy As a Humanistic Discipline*, ed. A. W. Moore (Princeton: Princeton University Press, 2006).

James Williams, 'Gilles Deleuze and A. W. Moore: Nonhuman Constructivism or Anthropocentric Narrative in Metaphysics', *Philosophical Topics*, Vol. 43, No. 1/2 (2015): 301-317.

Cary Wolfe, *Animal Rites: American Culture, the Discourse of Species, and Posthumanist Theory* (Chicago: University of Chicago Press, 2003).

Richard Wolin, ed., *The Heidegger Controversy* (Cambridge: MIT Press, 1993).

Allen W. Wood 'Kant's Compatibilism', in Allen W. Wood, ed., *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), pp. 73-101.

Mark Wrathall, '[Review of] *The History of Beyng*', *Notre Dame Philosophical Reviews* (2016): https://ndpr.nd.edu/reviews/the-history-of-beyng/

Unpublished papers:

Rainer Forst, 'Regression, Reason and Democracy', paper delivered at the "Philosophy and Social Science" Conference, University College Dublin, Saturday 19th May 2022.

Sean Kelly, 'The Proper Dignity of Human Being: Later Heidegger and the Philosophical Tradition', paper delivered at the Post-Kantian European Philosophy Seminar, University of Oxford, Tuesday 19th November 2019.