ARTICLE

Radical Evil, Social Contracts and the Idea of the Church in Kant

Jacqueline Mariña

Purdue University
Email: marinaj@purdue.edu

Abstract

In this article I argue that Kant's understanding of the universality of radical evil is best understood in the context of human sociality. Because we are inherently social beings, the nature of the human community we find ourselves in has a determinative influence on the sorts of persons we are, and the kinds of choices we can make. We always *begin* in evil. This does not vitiate responsibility, since through reflection we can become aware of our situation and envision ourselves as members of a different community, one with different expectations, making genuine virtue possible. This understanding of radical evil helps to make sense of Kant's high regard for the church in *Religion*.

Keywords: religion; radical evil; freedom; social contracts; church; unsocial sociability; ethical community; fundamental disposition; moral hope; anthropology

1. Introduction

It has often been pointed out that there seems to be an inconsistency in Kant's account of radical evil: Kant cannot have it both ways, he cannot, at one and the same time, claim that radical evil is *universal,* and that it is a *freely chosen* disposition (for instance, Quinn 1984). For if the individual is responsible for it, and if it is freely chosen, then it stands to reason that we cannot make any claims regarding the *universality* of radical evil. On the one hand, if radical evil is something freely chosen, then it is certainly possible that at least one individual might freely choose *not* to have a radically evil disposition. On the other hand, if a universal claim regarding radical evil is to have an adequate ground, then we would have to give an *a priori* account of why *all* human beings fall into radical evil. We would need some kind of formal proof for why human beings, one and all, are radically evil; this universality would have to be grounded in some kind of necessity. Such necessity, however, seems to vitiate freedom. There have been two basic approaches to this issue: one that strives to provide a kind of deduction for radical evil (e.g. Allison 1990; Morgan 2005; Palmquist 2008), and another that reduces Kant's claim to a kind of general universality. It is merely an anthropological claim, and as such, needs no proof going much further than the litany of woeful examples provided by Kant. This is basically Wood's approach (Muchnik 2010 also grounds accounts of radical evil – at least the

propensity – in grounds that can be investigated anthropologically). Importantly, Wood highlights the importance of the role of unsocial sociability in illuminating the human condition, a thesis first put forward in Anderson-Gold 1991, adopted by Wood (Wood 1999 and 2020: 76) and worked out further in Pasternack 2017. I argue that Kant's argument for the universality of radical evil attaches to his understanding of it as pertaining to the *species.* Hence, we cannot give an *a priori* account of radical evil as something that can be derived from the *concept* of the human being. But if we can give an account of how it attaches to *this* particular species, then we can see how we might ascribe radical evil to all members of *this* species of human beings, without falling into the mistaken belief that Kant held that radical evil follows inevitably from the concept of a rational but finite being of needs (Allison 1990: 154–7). My account is in some ways similar to that put forward in Firestone and Jacobs 2008, but differs significantly in that I do not emphasize any kind of Aristotelian account of secondary substance. Rather, I provide an account of radical evil that takes into account the *species* character of human beings, that is, the ineradicably social dimension of all human activity. This brings to the fore the importance of unsocial sociability in our understanding of the nature of radical evil, as well as illuminating the importance of Kant's discussion of the church as an antidote to unsocial sociability in books three and four of the *Religion.*

  While the analysis of radical evil in terms of unsocial sociability put forward by interpreters such as Anderson-Gold and Wood are on the right track, these analyses do not go far enough, or give a completely adequate account of why criticisms of this approach, such as those of Grenberg 2005 and Morgan 2005, are misguided. My own approach, sympathetic to that of Anderson-Gold, Wood and Pasternack, provides an analysis of Kant's thesis of the universality of radical evil in terms of social contracts. Through this approach we can further understand the importance of the social contract with the ethical community for both the change of heart and progress in virtue. This highlights the significance of conditions in the historical arena for the individual's capacity to progress in virtue. While the individual must certainly do all that is in her power to continue upon the path of virtue (*facere quod in se est*), there are issues having to do with *the others,* and the common projects that one must undertake with them, that are beyond the individual's control, in which the idea of the hope in, and actual striving *for*, the realization of the ethical community plays an important role. The moral individual must join an ethical community and do everything in her power (a) to act *as if* the ethical community is *in some* way a reality, and (b) work to ensure that the ways that persons relate to one another in an historically situated instance of an ethical community live up to the ideal. Joining an ethical community, and in some such way entering a social contract with it, is a condition of the possibility of the achievement of moral goals.

  This article will be divided into three parts. Section 2 provides an initial understanding of Kant's analysis of the universality of radical evil in terms of the *social condition,* or more precisely, the implicit and explicit social contracts that we find ourselves in. I then discuss these contracts in relation to the idea of unsocial sociability. Section 3 relates the social contracts we are in to both the fundamental disposition and to the propensity. I show why such social contracts, which are always already in place, condition all our choices such that we always *begin* in the condition of evil. This fact, however, does not contradict our freedom and personal responsibility, since it is our responsibility to enter into the kinds of community providing opportunities for genuine reciprocity. The kinds of social contracts we always already

find ourselves in, however, reveal why the particular human being, as a member of this particular race and its condition, begins in an evil condition and must work its way out of it. Lastly, in section 4, I very briefly discuss the role of the social contract in the ethical community and its relation to the proper attitude of moral hope.

2. Radical evil and social contracts

Kant begins his argument for the universality of radical evil in a somewhat puzzling fashion. In support of this thesis, what he offers are two empirical, anthropological observations concerning life in a state of nature and in civil society. That he does so has puzzled many of his readers, who have dismissed his arguments as muddled: he seems to be arguing for a strictly universal thesis through a couple of empirical observations. To some interpreters, such as Henry Allison (1990: 154–7), only a strictly *a priori* argument or deduction will suffice to bolster Kant's claim. Wood, on the other hand, correctly points to the fact that Kant is not here providing a deduction, but is rather making a claim based on the results of 'anthropological research' (Wood 2020: 86). As Kant notes:

> However, that by the 'human being' of whom we say that he is good or evil by nature we are entitled to understand not individuals (for otherwise one human being could be assumed to be by nature good and another evil, by nature), this can only be demonstrated later on, if it transpires from anthropological research that the grounds that justify us in attributing one of these two characters to a human being as innate are of such a nature that there is no cause for exempting anyone from it, and that the character therefore applies to the species. (6: 25–6)[1]

While Wood is correct in pointing to this fact, he is unable to show just how this anthropological research supports Kant's claim that there is 'no cause for exempting anyone' from radical evil, and 'that the character therefore applies to the species'. On what grounds could such anthropological research bolster such universal claims?

Importantly, Kant notes that the quality of radical evil cannot be inferred from the *concept* of (the human being's) species. Endeavours to provide a deduction of radical evil from the idea of a rational yet sensuously conditioned being are bound to fail. For it is entirely possible that there should be a species of rational finite, sensuously conditioned beings who do not have the quality of radical evil. Such beings would have two disparate sources of motivation, on the one hand the sensuously conditioned desires and on the other the desires stemming from pure practical reason, but they would *never* make satisfaction of the moral law conditional upon the satisfaction of the sensuously conditioned desires. As Kant notes, '"He is evil by nature" simply means that being evil applies to him considered in his species; not that this quality may be inferred from the concept of his species' (6: 32). It is this instantiation of a species of rational yet sensuously conditioned beings that is beset by radical evil, and as members of this community, radical evil can be attributed to each one of us. But what is it about being members of *this* particular instantiation of a species of rational yet finite and sensuously conditioned beings that leads to the inevitability of radical evil being true of each and every one of us, so that 'we may presuppose evil

as subjectively necessary in every human being, even the best' (6: 32)? The answer lies partially in the fact that the human being is what Marx had called a species being. She accomplishes nothing on her own, but only with the others with whom she stands in society. Kant, who mentions the views of both Hobbes and Rousseau (6: 33), was well aware of this. And if the human being is indeed such a species being, it stands to reason that the state of that society, and the way the human being is set to interact with its members, conditions the very nature of the projects she undertakes, as well as their possibilities for completion.

Just after Kant spares himself the 'formal proof' for the corrupt propensity, he mentions two social conditions, the state of nature and civil society. Examples of human evil abound in both. Against Rousseau, who remarked, regarding human beings in the state of nature, upon 'the calmness of their passions, and their ignorance of vice',[2] Kant calls attention to 'scenes of unprovoked cruelty' and 'vices of savagery' in the state of nature (6: 33). And in regard to human beings in their civilized state, in which they have entered the social contract and therefore the state of civilization, he details a 'long melancholy litany of charges against humankind', not the least of which is a 'secret falsity' in even the best of friendships, so that 'the misfortunes of our best friends' do not 'altogether displease us' (6: 33). It is clear from the rest of the book that Kant understands the social condition as a very significant factor in the propagation of radical evil. As he notes in book three, it is because of the '*lack of principle which unites them*' that human beings 'deviate through their dissensions from the common goal of goodness, as though they were instruments of evil' (6: 97; italics added). The established *social principles* of human interaction are of the utmost importance in the radicality of human evil.

If we take this suggestion seriously, we can see why Kant can argue for the universality of radical evil *without* this universality vitiating freedom, or without him attributing this universality to a necessity grounded in our finite and sensuously conditioned nature (on this point, Wood is correct to argue against Allison). And while we cannot provide a derivation of radical evil from the concept of the human being, we can note that because the human being is a *species* being, and because our species is at the state of development in which it is, with the kinds of social arrangements it currently has in place, all human beings who are members of this species at this point are beset by radical evil. Their being beset with this problem is in no small part due to the kinds of social contracts, both *implicit* and *explicit,* into which the human being finds him or herself as already having entered. These contracts govern what it is that is expected out of interactions with one another, as well as the *modes* of those interactions.

3. Social contracts, radical evil and responsibility

In *Religion,* Kant defines radical evil in terms of the subordination of incentives. As rational, but sensuously conditioned finite beings of needs, we have two kinds of desires or incentives for action: those connected with the lower faculty of desire, stemming from our embodiment, and relating to our happiness, and those stemming for the higher faculty of desire, namely our reason, through which we give ourselves the moral law. *Both* incentives are present in our maxims. What makes an individual good or evil is 'which of the two [incentives] he makes the condition of the other' (6: 36). The evil individual acts in accordance with morality *so long as it is convenient,*

but flouts the moral law when the cost of acting in accordance with it is too high. Here achieving one's happiness is made the condition of following the moral law. How does this account of radical evil dovetail with the social account? Grenberg (2005: 35), for instance, faults the social account (especially Wood's) for several reasons, one of which is that 'we cannot explain, in social terms alone, why it is that fear and anxiety develop in social situations'. But as Wood (2020: 78) notes, Kant clearly does have such a social account (for instance at 6: 93–4), so it is important to understand how the propensity to evil, which in an important sense is rooted in the individual and his or her maxim, can at the same time have a social and historical origin.

 An important passage from *Religion* clarifying the social and historical origins of radical evil is the following:

> The predisposition to humanity can be brought under the general title of a self-love which is physical and yet *involves comparison* (for which reason is required); that is, only in comparison with others does one judge oneself happy or unhappy. Out of this self-love originates the inclination *to gain worth in the opinion of others,* originally, of course, merely *equal worth:* not allowing anyone superiority over oneself, bound up with the constant anxiety that others might be striving for ascendancy; but from this arises gradually an unjust desire to acquire superiority for oneself over others. Upon this, namely upon *jealousy* and *rivalry,* can be grafted the greatest vices of secret or open hostility to all whom we consider alien to us. These vices, however, do not really issue from nature as their root but are rather inclinations, in the face of the anxious endeavor of others to attain a hateful superiority over us, to procure it for ourselves over them for the sake of security, as preventive measure; for nature itself wanted to use the idea of such a competitiveness (which in itself does not exclude reciprocal love) as only an incentive to culture. Hence the vices that are grafted upon this inclination can also be named vices of culture. (6: 27)

What counts as happiness is here developed in terms of our *sociality*. The good in question is *worth in the opinion of others.* The individual understands herself, counts herself happy and comes to know who she is through the other's opinion of her. She is happy just insofar as the other counts her as *having worth.* And because this question of the granting of worth also impinges upon our freedom (for the individual who grants, or is forced to grant, the other more worth than he or she herself has is in some kind of condition of slavery), the individual must be very careful in this granting of worth to the other. There is always the danger that her valuing of the other will not be reciprocated, or not fully. Hence dealings with the other are always fraught with danger. Here something along the lines of Hegel's master/slave dialectic, or Sartre's account of the battle between self and other when it comes to the 'the look', is in play. A battle for esteem ensues, for who counts more than the other, and regarding who, therefore, is to be granted privilege in the social arena. Because we are social beings, and because so much is at stake (one's own freedom!) one oversteps one's bounds and attempts to dominate and control the other. But this is, precisely, a kind of *not* recognizing the other as an end in herself, a taking of her principally as a kind of wild force that must be controlled. And so morality is made conditional upon what one takes to count as one's happiness: I will be moral so long as my freedom is not

threatened in any way, so long as *I* am in control, as long as *I* can extract from you an acknowledgement of my greater worth.

We can see here the outlines of the relation between the inversion of incentives and the social, historical and political context of radical evil. Important here is the fact that sociality is *essential* to our humanity; it is not something tacked on to our nature as individuals. The predisposition to humanity, in which we 'strive to gain worth in the opinion of others', is one of the *essential* predispositions of the human being (6: 27). It would therefore be misleading to take too seriously Kant's remark that the human being can count herself happy 'insofar as [s]he exists in isolation', but only *becomes* unhappy when she begins to compare herself with the others (6: 93). Kant is very well aware of the fact that the human being is essentially social.

Because we are social beings, we are always already intrinsically wired to care about how other people think of us. Note that this battle for esteem – its existence and how it plays itself out – is going to depend heavily on the *rules* by which one understands everyone else to be playing. Because *they* are playing by such rules, and I depend upon them when I act in concert with them (which I always must, since I depend on the others for almost everything), I have to act in accordance with such rules too. Not doing so puts me at immense risk and threatens my life and happiness. Not just in terms of my ability to get by in the world, although that too, can be threatened, but also in terms of the threat of loss of one of the *primary* goods that makes life meaningful to us to begin with, namely, the esteem of others.

There are, then, promising avenues for relating Kant's understanding of the inversion of incentives to the social and historical origin of radical evil.[3] One objection to understanding radical evil in these social and political terms is that, as Wood (2020: 78) puts it, 'to place evil in the context of human competitiveness is to exempt human individuals from responsibility for it'. This objection has been developed by Grenberg. She notes that 'the story of the social conditions of evil that Wood gives could still be construed as large forces outside the agent and impinging on her which undermine individual responsibility for her choice. ... our own culpability has been undermined' (Grenberg 2005: 35–6). Wood has countered this objection by noting that it depends on the mistaken notion that for Kant 'the social condition must consist of nothing but external causal interactions in the world of appearance, while free action must take place somewhere else entirely ("up in the noumenal world")' (Wood 2020: 78). But it is not at all clear that this is really the fundamental issue. The question is whether, if the conditions of the social world are *sufficient* to account for our radical evil, there is any real sense in which we could say that we *could have avoided it,* so that we can count ourselves responsible in some significant sense.[4] And this once again brings us back to the problem of how we can count ourselves responsible for radical evil, that is, count ourselves as *freely* espousing it, if it is something that is universal, and therefore necessary, so that we cannot escape it. One way of tackling this problem is to admit that we are, in fact, born into social situations where the rules of the game make our corruption inevitable *given* the kinds of social beings that we are. And it may be that the individual is never strong enough, *on her own*, to fully resist radical evil – that is, Walpole's thesis may be right. But this does not mean that *upon reflection* the individual cannot recognize that he or she is under an obligation to join a different kind of community (the ethical community) in order to be able to cope more effectively with those kinds of pressures. It is this obligation and the genuine capacity to

join such a community that undergirds our genuine freedom and responsibility, for here the individual does have choice.

4. Social contracts and the idea of the church

Wood, Gold and Pasternack are correct to insist that Kantian ethics is not individualistic but communal. This is in part because realization of all human goals in some way depends on 'the others' who surround us. Hence the nature of the principle uniting human beings with one another – for example, whether that principle is reflective of trust and mutuality of recognition, or of fear and competitiveness – plays a major role in whether the human being is subject to radical evil. Given the social character of human beings, and given the weakness of the single individual, radical evil is something that can never be eradicated unless the individual is a member of the ethical community, where the public common principles of action are devoted to the genuine recognition of the other as an end in him or herself. As Pasternack (2017: 454) notes, the ethical community 'depends on a way of being together not shaped by unsocial sociability'. In this community, 'you will succeed through my success, and vice-versa'. Such a community guides itself by the principles of virtue. Since these, however, have to do with the inner motive, which cannot be coerced or in fact really known publicly, the individual who joins the church joins an invisible kingdom under a common head, namely God, and so obeys the moral law *as if* it were a divine command. The ethical society is an invisible kingdom in which it is the *inner principle* of the heart that counts (6: 102). Only God knows the heart, and only God can establish a situation, presaged in the church, in which this inner principle *itself* has the power to bring forth good, since the good in question just is the mutuality of recognition, namely the practical love of the members of the community for one another. And this good can only be brought about through the faith, or trust, that this new situation can indeed be brought about.

Wood (2020: 171) tells us that for Kant 'the historical origin of religion ... was an attitude of slavish submission to supernatural powers', and the church itself had its origin there. So for Wood, the church, like all other human institutions, begins as a thing far removed from the aims of reason, and must develop. Here Wood quotes Nietzsche's *Genealogy*, 'all good things were formerly bad things', and claims that this is a good way of describing Kant's philosophy of history, as well as his understanding of the origins of the church. But this is not Kant's position. Kant asks 'how could one expect to construct something completely straight from such crooked wood?' and answers, 'To found a moral people of God, is therefore, a work whose execution cannot be hoped for from human beings but only from God himself' (6: 101); 'God must himself be the author of this Kingdom' (6: 152). And if we are to take Kant at his word, then there is certainly something not quite right about Wood's account of Kant's position. The good does not just somehow develop out of the bad. Kant's picture, instead, is of some kind of revelation or intervention. The teacher of the gospel is 'the founder of the first true church', and it is he who awakens what is 'inscribed in the heart of all human beings' (6: 159). It may very well be that such a teacher *can* found the moral community because Walpole's thesis is not true of him, and he alone expresses the moral law in its purity. And in doing so,

he opened the doors of freedom to all who, like him, choose to die to every-thing that holds them fettered to earthly life to the detriment of morality; and among these he gathers unto himself 'a people for his possession, zealous of good works,' under his dominion, while he abandons to their fate all those who prefer moral servitude. (6: 82)

The origin of the ethical community must be pure; out of evil nothing good can arise, and good arises if it is only already present, perhaps only as a seed, and needs only to be awakened. But its awakening had to come from a divine source. Contra Wood's thesis, the church itself does not *arise* out of our unsocial sociability.

Recognition of the social dimension of radical evil, its central importance in under-standing human nature, also opens the way for understanding the centrality of the ethical community – the church – in Kant's *Religion.* I have outlined the ways in which a further exploration of radical evil as grounded in our social nature can explain the universality of radical evil while at the same time allowing for the fact that finite sensuously conditioned rational agents do not necessarily fall into it. Its universality is due to the weakness of human nature combined with the kinds of societies we find ourselves born into. On this understanding we can see why Kant argued that the idea of the church presents us with the most significant object of moral hope in combating radical evil. Lastly, it is significant that the ethical community is *what* it is because we take it to have been founded by God *through* his representative, his Son:

It could best of all be likened to the constitution of a household (a family) under a common though invisible moral father, whose holy son, who knows the father's will and yet stands in blood relation with all the members of the family, takes his father's place by making the other members better acquainted with his will; these therefore honor the father in him and thus enter into a free universal and enduring union of hearts. (6: 102)

The ethical community is established through being bound in fellowship with the Son of God. For only one who has completely died to the values of 'the world' can establish such a community on a completely new footing, and it is this new principle and the society established through it that can join together both the moral command *and* the aim of happiness, for here happiness consists in the mutual regard that members of the ethical community grant one another.

Notes

1 All citations from Kant's *Religion* (tr. George di Giovanni) are from Kant 1996. The Academy edition pagination is provided after each citation.
2 Jean-Jacques Rousseau, *Discourse on the Origin and Basis of Inequality among Men* (1754), in Rousseau 2002: 106.
3 Grenberg tries to provide such an account, stressing the finite rational agent's *dependence*. Hence she notes that 'dependent agents seek a happiness based in part upon meeting the needs inherent in their dependent natures. Dependent beings cannot deny having reason to seek out persons and things in the world who and which meet their needs and wants' (Grenberg 2005: 37). And she continues, 'The depen-dent agent seeking to love herself properly is forced to admit the loss of the dream of perfect happiness. And it is only in the face of this inevitable loss that we can understand the anxiety of individual finite agents not just in social situations, but more generally. Dependent agents in this unstable condition tend

not just to "assert" their right to happiness, but to get anxious about further loss, and thus to *over*-assert their right to happiness' (p. 38). But Grenberg fails to account for how radical evil is *itself* grounded in the fact that sociality is essential to our humanity. It is not that we depend on others because we are finite and have survival needs, and therefore need them for those purposes, and that *this* is the ground of our anxiety. It's that at our very core we depend upon the others for who we understand ourselves to be; their esteeming us is a primary good for us, independently of whether that esteem is tied to our capacity to provide for our physical well-being. Grenberg fails to see this.

4 One helpful way of clarifying these issues is this: while our sociality is a *necessary condition* of radical evil, it is still an open question whether it is a *sufficient* condition of it. Recognizing our sociality as a necessary condition of radical evil is one way of emphasizing the importance of the social origins of radical evil without thereby affirming that the social character of the species is *sufficient* to account for an agent's falling into radical evil. Only if the latter is affirmed does this vitiate the personal responsibility of the individual. Now, it is important to note that if we affirm that it is the individual's free choice to engage with the other in a particular way, we are still left with the problem of how to account for the *universality* of radical evil. Common sense tells us that, given the kinds of finite and dependent beings that we are, our social condition inclines but does not necessitate us to play by certain rules, but that this inclining is very strong.

References

Allison, Henry (1990) *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.

Anderson-Gold, Sharon (1991) 'God and Community: An Inquiry into the Religious Implications of the Highest Good'. In Philip Rossi and Michael Wreen (eds), *Kant's Philosophy of Religion Reconsidered* (Bloomington, IN: Indiana University Press), 113–31.

Firestone, Chris and Nathan Jacobs (2008) *In Defense of Kant's Religion.* Bloomington, IN: Indiana University Press.

Grenberg, Jeanine (2005) *Kant and the Ethics of Humility*. Cambridge: Cambridge University Press.

Kant, Immanuel (1996), *Religion and Rational Theology*. Trans. and ed. Allen W. Wood and George di Giovanni. Cambridge: Cambridge University Press.

Morgan, Seiriol (2005) 'The Missing Formal Proof of Humanity's Radical Evil in Kant's "Religion"'. *The Philosophical Review*, 114(1), 63–114.

Muchnik, Pablo (2010) 'An Alternative Proof of the Universal Propensity to Evil'. In Sharon Anderson-Gold and Pablo Muchnik (eds), *Kant's Anatomy of Evil* (Cambridge: Cambridge University Press), 116–43.

Palmquist, Stephen (2008) 'Kant's Quasi-Transcendental Argument for a Necessary and Universal Evil Propensity in Human Nature'. *Southern Journal of Philosophy*, 46(2), 261–97.

Pasternack, Lawrence (2014) *Kant's Religion within the Boundaries of Mere Reason: An Interpretation and Defense.* London: Routledge.

—— (2017) 'Restoring Kant's Conception of the Highest Good'. *Journal of the History of Philosophy*, 55(3), 435–68.

Quinn, Philip (1984) 'Original Sin, Radical Evil, and Moral Identity'. *Faith and Philosophy*, 1(2), 188–202.

Rousseau, Jean-Jacques (2002) *The Social Contract and the First and Second Discourses.* Trans. and ed. Susan Dunn. New Haven: Yale University Press.

Wood, Allen W. (1999) *Kant's Ethical Thought.* Cambridge: Cambridge University Press.

—— (2020) *Kant and Religion.* Oxford: Oxford University Press.