

Daniel Eggers, Religious conflict and moral consensus: Hobbes, Rawls, and two types of moral justification

This is the peer-reviewed manuscript of the book chapter “Religious conflict and moral consensus: Hobbes, Rawls, and two types of moral justification”, published in Laurens van Apeldoorn and Robin Douglass (eds.), “Hobbes on Politics and Religion”, Oxford: Oxford University Press, 2018, pp. 239 – 255. The text is reproduced by permission of Oxford University Press (<https://global.oup.com>). The final version is available at: <https://doi.org/10.1093/oso/9780198803409.003.0015>.

The claim that Hobbes’s political theory was shaped by the experience of the English civil war and the political and religious conflicts that preceded it is arguably the main common place about his philosophy. Both within and without Hobbes scholarship more narrowly construed, the idea that the appearance of his theory is the result of the mixture of political and religious strife witnessed during his lifetime has countless times been repeated and hardly ever been challenged.

Yet, there are different ways in which the experience of the political and religious struggles shaped Hobbes’s theory, as there are obviously quite different strategies by which one can respond to such struggles, and to the challenge posed by religious pluralism in particular. One strategy was for Hobbes to directly engage in theological discussion and provide his own interpretations of the key tenets of Christian religion. A related strategy was to provide scriptural interpretations and to demonstrate that his moral and political teachings are compatible with the teachings of the Bible. As many Hobbes scholars have emphasized, there is a lot of evidence that he pursued both these strategies. We find in all his major political works explicit discussions of theological concepts and explicit efforts to identify his moral and political doctrine as one that any true believer can accept.

However, it is natural to think that the experience of religious struggle should also have shaped Hobbes’s political doctrine, and the moral theory underlying it, in a more direct manner. Given that the experience characteristic for Hobbes and his time was that of the *conflict* between different groups of citizens and their diverging religious convictions, we might expect him to have tried to devise a moral and political theory that is equally acceptable to readers of diverse religious and ideological backgrounds.

That Hobbes pursued this third and somewhat autonomous strategy is an idea we find in Hobbes scholarship, too. What makes this strategy especially interesting is that it has been

taken up by modern Hobbesians such as David Gauthier and Peter Stemmer, who develop their moral theories with the explicit aim of overcoming problems of religiously motivated moral disagreement in the modern world.¹ As commentators have sometimes pointed out, what is characteristic of Hobbes's political argument is that it starts from relatively modest and widely acceptable anthropological assumptions, most notably the assumption that human beings have a strong or even overriding interest in their own preservation (at least insofar as they are rational).² It then ventures to show that quite considerable normative constraints on human behaviour can be derived from these assumptions, constraints that must rationally be accepted by anyone who accepts the original starting point. It is this idea which we find exemplified in Gauthier's and Stemmer's theories and which they present as a crucial element of their Hobbesian heritage.³

In the wake of Rawls's *Political Liberalism*, it has become customary to describe consensualist strategies in moral and political philosophy as the attempt to found an overlapping consensus between different conceptions of the good. However, the general strategy of appealing to a moral consensus that can ground one's political argument in the face of religious pluralism is not confined to the context of Rawls's theory of justice and should therefore not be confused with Rawls's particular version of it. My aim in this paper is to analyse the way in which Hobbes pursues a consensualist approach and to contrast it with Rawls's quite different effort. My reasons for choosing Rawls's theory as a foil of comparison are threefold. First, as just indicated, the notion of a moral consensus is more closely associated with Rawls than with any other philosopher in the current philosophical debate; secondly, like Hobbes, Rawls attempts to establish such a consensus within the general philosophical framework of contract theory; thirdly, Rawls pursues a consensualist approach that is interestingly and importantly different from Hobbes's approach in that it

¹ See David Gauthier, *Morals by Agreement* (Oxford: Oxford University Press, 1986); and Peter Stemmer, *Handeln zugunsten anderer* (Berlin and New York: de Gruyter, 2000). Stemmer explicitly emphasizes that people's diverging ethical views are largely the result of different religious assumptions (p. 204) and that, in response to this, his objective is one of devising a moral theory that applies to all rational human beings in the same way (p. 210).

² See, for example, Konrad Ott, *Moralbegründungen zur Einführung* (Hamburg: Junius, 2001), p. 123f. The view is also implicitly suggested by Gauthier, *Morals*, p. 17; and Robert Sugden, 'Rationality and Impartiality: Is the Contractarian Enterprise Possible?', in *Rationality, Justice and the Social Contract: Themes from Morals by Agreement*, ed. David Gauthier and Robert Sugden (New York and London: Harvester Wheatsheaf, 1993), p. 157f.

³ See, in particular, Gauthier, *Morals*, p. 17.

relies on a different idea of moral justification. Since Rawls's theory, therefore, demonstrates that (and how) an appeal to moral consensus different from the one pursued by Hobbes is possible, and possible within a broadly contractarian framework, it can help us both to systematically classify Hobbes's approach and to identify its possible weaknesses.

In section 1, I first provide textual evidence that Hobbes's moral theory appeals to a *consensus omnium* and then systematically classify his argument as an instance of what I refer to as extra-moral justification. In section 2, in contrast, I briefly describe Rawls's theory as an instance of intra-moral justification. In the third section, I defend this way of contrasting Hobbes's and Rawls's justificatory strategies against an objection that might be raised on the basis of Lloyd's recent interpretation of Hobbes's moral theory. In the final section, I critically assess the potential of the Hobbesian strategy for grounding a moral consensus under conditions of religious diversity. My tentative conclusion will be that the possible philosophical advantages of the fundamental justification provided by Hobbes and modern Hobbesians do not make up for the losses that a Hobbesian moral theory incurs in terms of compatibility with common sense morality. While Rawls's appeal to a moral consensus may face problems in its own right, the more modest idea of moral justification on which it relies, and which is explicitly rejected by Gauthier and Stemmer, provides no reasons for thinking that the quite different Hobbesian appeal must ultimately yield better results.

I.

In all his major political works Hobbes's political argument begins with a description of the state of nature that is meant to provide the basis for his political conclusions by showing that no human being can rationally desire to live in state that lacks the characteristic structures of civil society. This account of the natural condition of mankind is in various ways designed to be able to claim wide or even universal acceptance. Hobbes's ultimate argument for why individuals need to leave the state of nature is that it is a state of war and hence ultimately incompatible with the goal of self-preservation.⁴ That this aim is one we can attribute to all

⁴ It has been suggested that Hobbes presents peace as intrinsically good and that the requirement to leave the state of nature might therefore be independent of any desire to preserve oneself. See, for instance, Bernd Ludwig, *Die Wiederentdeckung des Epikureischen Naturrechtes* (Frankfurt am Main: Klostermann, 1998), p. 374. However, there is ample evidence that Hobbes takes the value of peace to be an instrumental value that consists in what peace contributes to the shared but individual aim of personal survival. See, for example,

human beings, at least in so far as they are rational, is repeatedly claimed in all of Hobbes's works.⁵ Moreover, Hobbes infers from the pivotal role assigned to the goal of self-preservation that his state of nature argument is one that no reader can reasonably reject. According to Hobbes, nobody can deny that the state of nature necessarily turns into a state of war, nobody will view such a state as a good for himself, and everybody will consider it advisable to leave the state of nature.⁶

The derivation of the state of war itself is also laid out in a way that is meant to ensure the assent of different groups of readers. It developed out of a handful of empirical assumptions about the physical and psychological features of human beings and their natural living conditions that are presented by Hobbes as beyond dispute. His initial assumption is that human beings are relevantly equal to one another with regard to their physical and mental capacities. Though Hobbes admits that there are considerable differences in individual abilities, he argues that most human beings are sufficiently similar to serve as a threat to one another and concludes from this everyone is well advised to conceive of others as equal and treat them accordingly.

Kinch Hoekstra has recently suggested that, instead of being derived from the equality of men's natural abilities, the requirement to acknowledge others as equal logically precedes and, in fact, grounds Hobbes's claims about natural equality.⁷ However, Hoekstra's account ultimately collapses into the standard position he strives to reject. According to Hoekstra, the reason why we should acknowledge others as equal is that this is necessary for peace because failing to do so would increase the probability of violent conflict. However, if—as we have already seen—peace is introduced by Hobbes not as an end in itself but only as a necessary means to self-preservation, then it seems that the requirement to acknowledge others as equal

Hobbes's claim that 'reason dictates to every man *for his own good* to seek after peace'. EL, 15.2 (my emphasis); see also L, 15, p. 242. Note also that the requirement to seek peace is explicitly qualified and applies only where individuals can do so without putting their preservation at risk (see L, 14, p. 198). Moreover, in describing the duties of the sovereign, which are grounded in the laws of nature, Hobbes does not refer to the goal of peace, but to the 'Preservation' of the subjects and 'other Contentments' (L, 30 p. 520) of their lives.

⁵ See EL, 14.6, 14.12, 14.13, 17.14; DCv, 1.7, 2.3. Hobbes goes as far as to describe the desire to avoid violent death as one of the two most certain postulates of human nature, DCv, 'Epistle Dedicatory'.

⁶ See DCv, 1.12, 3.31. See also Hobbes's remarks that the state of war is a state which all men by nature abhor and that all men want to leave the state of nature once they realise the evils that necessarily attach to it, DCv, 'Epistle Dedicatory'.

⁷ See Kinch Hoekstra, 'Hobbesian Equality', in *Hobbes Today: Insights for the 21st Century*, ed. S. A. Lloyd (Cambridge: Cambridge University Press, 2012), p. 112.

cannot ground Hobbes's claim about natural equality because it presupposes it. If it were not for relevant natural equality and the fact that 'the weakest can kill the strongest', individuals would simply have insufficient reasons to fear and avoid violent confrontation.

Hobbes's assumption of relevant equality is then hooked up with the claim that certain internal and external conditions of human life in the state of nature actually lead the individuals into conflict with one another. As I have tried to demonstrate elsewhere,⁸ Hobbes's description of the two main sources of conflict—competition and anticipation—is dependent on the relatively modest and defensible presupposition that the state of nature is a condition of uncertainty and moderate scarcity. His description of the third source of conflict—glory—is bound to be slightly more controversial. It needs to be emphasized, however, that Hobbes does not allege an active disposition to extort respect from others wherever this is possible, but the more passive disposition to defend one's reputation *if* it is challenged. Moreover, though Hobbes suggests that this disposition is one every individual possesses, he leaves open whether it is equally strong in everybody. It is also important to note that, in order for his ultimate conclusion to go through, Hobbes does not need to put much emphasis on the third source of conflict in the first place: Hobbes explicitly describes the state of war as consisting, not in constant actual fighting, but only 'in the known disposition thereto, during all the time there is no assurance to the contrary',⁹ and the first two sources are sufficient to turn the natural state into a state of this kind.

To say that Hobbes may legitimately view his derivation of the state of war as resting on an uncontroversial anthropological foundation, however, is not to say that the message was well received by his readers. It is this part of the argument that, up to this day, has led people to criticize Hobbes for expounding an overly negative view of human nature. Yet, the way Hobbes responds to this objection only goes to confirm his interest in appealing to a *consensus omnium*. In the second edition of *De Cive*, Hobbes tries to dispel the worry that his argument presents human beings as evil by nature and to identify his anthropological assumptions as ones which his readers accept for themselves—if not by word, then by deed. Hobbes reminds his readers of the fact that, though living in the relative safety of civil society, they usually do not travel without a weapon or go to bed without locking their doors, just as commonwealths defend their borders with garrisons of soldiers, even when they are at

⁸ See Daniel Eggers, 'Hobbes and Game Theory Revisited', *Southern Journal of Philosophy* 49, no. 3 (2011), p. 199ff.

⁹ L, 13 p. 192.

peace with their neighbours—which would all be irrational if one were to deny that certain aspects of human psychology or human social life make violent conflict probable.¹⁰ We find a further response in the fact that Hobbes significantly revises his derivation of the state of war in his later works. The text of *Leviathan*, on which I relied above, is already the result of some important modifications in Hobbes's description of the several sources of conflict which all tend to stress the role of the rational pursuit of self-preservation over the role of anti-social human inclinations.

That Hobbes is interested in securing the uncontroversial character of his argument cannot only be seen from the way he describes and revises his derivation of the state of war, but also from the way in which he sets up his doctrine of natural right and natural law. In justifying the initial right to self-preservation in the *Elements*, Hobbes claims that it is not against reason that an individual should try to defend his life and limbs and adds: 'that which is not against reason, men call RIGHT'.¹¹ Moreover, even after arriving at the radical conclusion that every man has a natural right to all things, Hobbes attempts to class his position as part of (philosophical) common sense, by adding '[a]nd for this cause it is rightly said: *Natura dedit omnia omnibus...*'¹² We find almost identical statements in the corresponding passages of *De Cive*,¹³ and in *Leviathan* we have a similar appeal to common sense in Hobbes's discussion of the necessity of anticipatory violence, where Hobbes remarks that if an individual strikes out pre-emptively against others, 'this is no more than his own conservation requireth, and is generally allowed.'¹⁴

As far as the doctrine of the laws of nature is concerned, the appeal to a universal consensus is most visible in Hobbes's more general descriptions of his own doctrine. In the 'Epistle Dedicatory' of the *Elements*, Hobbes alludes to previous controversies about morality and justice and affirms his intention to build the laws of nature, 'which hitherto have been built in the air', on safe and uncontroversial principles 'till the whole be inexpugnable'.¹⁵ That the laws of nature he subsequently formulates actually meet this standard is argued in the concluding passages of chapter seventeen. Hobbes emphasizes that every individual must call good what contributes to his preservation and must therefore also

¹⁰ See L, 13 p. 194.

¹¹ EL, 14.6.

¹² EL, 14.10

¹³ See DCv, 1.7, 1.10, 2.1.

¹⁴ L, 13, p. 190.

¹⁵ EL, 'Epistle Dedicatory'.

approve of the ways of behaviour laid out by the laws of nature. According to Hobbes, the actions commanded or forbidden by the laws of nature are ‘that good and evil, which not every man in passion calleth so, but all men by reason’.¹⁶ Again, we find very much the same claims in *De Cive* and in *Leviathan* where emphasizes that ‘all men agree on this, that Peace is Good, and therefore also the way, or means of Peace, which (as I have shewed before) are *Justice, Gratitude, Modesty, Equity, Mercy, & the rest of the Laws of Nature, are good*’.¹⁷

As with his description of the various sources of conflict in the state of nature, Hobbes’s attempt to identify his doctrine of natural right and natural law as one that is fully in line with what people ordinarily think was hardly successful with regard to his contemporary readers. Yet, since at least the late 1960s, there has been a tendency in Hobbes scholarship to assess Hobbes’s discussion of natural right and natural law much more charitably. Moreover, as already indicated above, modern contractarians like Gauthier and Stemmer have taken up Hobbes’s argumentative strategy and have done so just because of the moderate and uncontroversial assumptions on which this strategy rests.

If we try to generally characterize the kind of moral justification that Hobbes provides us with, then the justification of the various laws of nature is a form of extra-moral justification in that Hobbes attempts to derive the moral necessity of certain forms of behaviour from wholly non-moral premises. The purpose of Hobbes’s discussion of morality is not to merely apply given moral principles to particular situations or to specific challenges that arise as part of human life, nor is it to justify some key moral principles by deriving them from a more general moral principle or from the moral standpoint as such. Rather, Hobbes wants to demonstrate that it is rationally necessary to act in accordance with certain moral principles without already drawing on any genuinely moral perspective or any genuine moral assumptions, such as those grounded in certain religious views of his time.

Though Hobbes himself may not describe his methodological approach in exactly these terms, there is sufficient evidence that he, like modern Hobbesians,¹⁸ conceives of it in this way and that his reliance on a form of extra-moral justification results from a conscious decision on his part. He explicitly describes the state of nature argument which grounds his moral theory as an ‘Inference, made from the Passions’ and thus as a piece of descriptive anthropology.¹⁹ In accordance with this, both Hobbes’s claim that human beings are

¹⁶ EL, 17.14.

¹⁷ L, 15, p. 242. See also DCv, ‘Epistle Dedicatory’, 3.31f.

¹⁸ See, in particular, Gauthier, *Morals*, pp. 4, 9f.

¹⁹ L, 13, p. 194.

relevantly similar in their physical and mental abilities and his various claims about the internal and external sources of conflict are descriptive claims that do not yet contain any commitment to particular moral principles or even as much as a moral standpoint. The same applies to his treatment of the fundamental desire for self-preservation from which Hobbes's laws of nature derive their normative authority.

II

Compared to the preceding discussion of Hobbes's moral theory, my treatment of Rawls will be relatively brief. The main reason for this is that there is no real need to verify the consensualist credentials of Rawls's theory: That Rawls aims at an overlapping consensus is not only explicitly emphasized by him, but also unanimously acknowledged among Rawls scholars. Rather than trying to substantiate a consensualist interpretation of Rawls's theory, then, we can focus on the specific nature of the Rawlsian consensus and on his justification of it.

The idea of an overlapping consensus is developed in a 1987 paper with the very same title and plays a crucial role in the argument of *Political Liberalism*, originally published in 1993. Rawls takes up the idea of an overlapping consensus because of certain problems with the original version of his theory.²⁰ However, the fact that Rawls develops the idea of an overlapping consensus only after the publication of *A Theory of Justice* should not mislead us into thinking that the idea of a moral consensus plays no role in his original theory. The difference between *A Theory of Justice* and *Political Liberalism* is that the former relies on the optimistic view that all or nearly all citizens may agree on one reasonable comprehensive doctrine of justice, while the latter admits that agreement will extend only to a political conception of justice more narrowly construed. What distinguishes the two theories, then, is not that one conceives of the acceptance of the principles of justice in terms of an underlying moral consensus while the other does not. It is rather that the consensus in question is conceived of differently. Accordingly, there is talk of consensus, and even of overlapping consensus, in *A Theory of Justice*,²¹ and in his introduction, Rawls even sets out his programme in a way that is reminiscent of the kind of consensualist strategy I have

²⁰ For Rawls's own account of these problems, see John Rawls, *Political Liberalism* (New York: Columbia Press, 2005), pp. xv–xviii.

²¹ See, for example, John Rawls, *A Theory of Justice* (Cambridge, MA: Belknap Press of Harvard University Press, 1971), pp. 387f, 581f.

attributed to Hobbes.²² For the purposes of this paper, I will therefore understand the idea of a moral consensus in a somewhat broader sense that can encompass the efforts made by Rawls in *A Theory of Justice*.

The general purpose of Rawls's theory is to formulate principles of justice for the basic structure of a well-ordered society. The focus of the justification of Rawls's two principles is provided by the question of why we can expect citizens with diverging comprehensive theories of the good to equally accept these two principles and, in particular, the list of basic liberties which complements them. The justification of *A Theory of Justice* draws heavily on the idea that there are such things as all-purpose means, i.e. goods which any individual having a theory of the good or a rational plan of life will necessarily desire. This idea lies at the heart of Rawls's theory of 'primary social goods' in its original version.²³ Primary goods, as introduced in *A Theory of Justice*, are 'things it is supposed a rational man wants whatever else he wants',²⁴ and the examples given by Rawls include rights, liberties, powers, opportunities, income, wealth, and the social bases of self-respect. His theory of the primary goods represents an important element of the argument from the original position: Given the veil of ignorance, which denies the parties any knowledge of their personal features or talents and their particular theories of the good, it is only in virtue of the presupposed desire for primary goods that the parties are in a position to choose any specific principles of justice at all.

In contrast, the justification of *Political Liberalism* crucially relies on two ideas that linger in the background of *A Theory of Justice*: the idea of society as a fair system of cooperation and the idea of the person. A person, according to Rawls, is 'someone who can be a citizen, that is, a normal and fully cooperating member of society over a complete life.'²⁵ In order to be a normal and fully cooperating member of society, an individual needs to have what Rawls refers to as the two 'moral powers'²⁶: a capacity for a sense of justice and a capacity for a conception of the good. According to Rawls, it is in virtue of these two powers that we conceive of persons as free and rational. Moreover, having these powers to the requisite degree in order to be a fully cooperating member of society is what makes people equal in the relevant political sense.

²² Ibid., p. 18.

²³ Ibid., p. 90.

²⁴ Ibid., p. 92.

²⁵ Rawls, *Liberalism*, p. 18.

²⁶ Ibid., p. 19.

Rawls's conception of the two moral powers introduces an additional criterion that guides the parties in the original position.²⁷ In Rawls's view, we can legitimately attribute to all citizens a higher-order interest in the development and full exercise of their two moral powers, given that these powers are what makes them fully cooperating members of society and free and equal persons. As the Rawls of *Political Liberalism* claims, it is this higher-order interest in the development and full exercise of the two moral powers that initially explains the citizens' interest in the various primary goods and that also allows them to weigh these goods against one another.

As with Hobbes, there are certainly different ways to characterize the type of moral justification we encounter in Rawls's argument. The point that I would like to focus on is that Rawls does not provide an extra-moral justification, but only what we may describe as an intra-moral justification. It is important to recognize that this is true of Rawls's argument in both its original and its revised form. The desire for primary goods is *prima facie* not so different from the kind of self-interest on which Hobbes's argument relies. However, given the veil of ignorance, which ensures that the decision made by the parties in the original position is not affected by any differences in the individuals' natural assets or their particular positions in society, Rawls's argument is by no means morally neutral. In its original form, therefore, Rawls's justification appeals to a consensus in two ways: by positively basing the choice of the principles of justice on a shared thin conception of the good (represented by the universal desire for primary goods); and by negatively excluding contingent differences in power or natural talent from providing reasons for choice because this would violate our considered judgements about fairness. While the former consensus may entirely be constituted by non-moral assumptions, the latter is clearly not.

In the revised argument in *Political Liberalism*, the moral basis for the justification of the two principles of justice is even more visible. The idea of the person and the idea of the two moral powers are just as morally tinted as the fundamental idea of society as a fair system of cooperation. Accordingly, the universal interests that guide the choice of the parties in the original position are not non-moral interests, but presuppose the recognition of general moral principles of reciprocity or fairness. This is also emphasized by the fact that the consensus Rawls describes is one between *reasonable* comprehensive doctrines, not one between rational comprehensive doctrines. According to the explicit distinction offered in *Political Liberalism*, individuals act reasonably, as opposed to merely rationally, if they

²⁷ For the following, see *Ibid.*, pp. 75f, 178f.

accept fundamental ideas of fairness and reciprocity and ‘desire for its own sake a social world in which they, as free and equal, can cooperate with others on terms all can accept.’²⁸

III.

The way Rawls appeals to a moral consensus, therefore, appears to be decidedly different from Hobbes’s way. However, before assessing the respective merits of the two types of moral justification, I would like to take up the possible objection that Hobbes is in fact much more of a Rawlsian than I allow. According to the ‘reciprocity interpretation of Hobbes’s moral philosophy’ recently advocated by S. A. Lloyd, Hobbes’s moral theory appeals to a Rawlsian notion of reciprocity in a way that makes it an instance of intra-moral rather than extra-moral justification.²⁹ In Lloyd’s view, Hobbes’s doctrine of the laws of nature is neither based on the non-moral desire for self-preservation nor on considerations of self-interest more generally. Rather, Hobbes’s point of departure is a conception of ourselves as rational agents and persons, according to which a person is somebody who is generally willing and able to offer justifying reasons for his behaviour. As Lloyd claims, the central idea of Hobbes’s moral theory is then provided by a ‘reciprocity theorem’, according to which offering certain considerations as justifying reasons for one’s actions commits oneself to accepting the same considerations as justifying the like actions of others.³⁰ Hobbes thereby provides ‘an early articulation and defense’³¹ of the idea of reasonableness we find in Rawls, and in particular ‘an insistence on the irreducibility of ... the reasonable to the rational.’³²

Lloyd is not always entirely clear on whether she understands the reciprocity theorem as a substantial moral principle demanding fairness and impartiality or merely in terms of the kind of conceptual principle usually referred to as the principle of universalizability. However, both the analogy with Rawls and her emphasis on Hobbes’s use of the Golden Rule

²⁸ Ibid., p. 50.

²⁹ S. A. Lloyd, *Morality in the Philosophy of Thomas Hobbes* (Cambridge and New York: Cambridge University Press, 2009), p. 7.

³⁰ Ibid., pp. 4, 220.

³¹ Ibid., p. xvi.

³² Ibid., p. 226f. It deserves to be noted that Rawls himself does not conceive of Hobbes’s theory in this way. According to Rawls, Hobbes explains the reasonable wholly in terms of the rational and, as a result, has no place for any moral rights or duties in the ordinary sense of these words. See John Rawls, *Lectures on the History of Political Philosophy*, ed. Samuel Freeman (Cambridge, MA: Belknap Press of Harvard University Press, 2007), p. 54ff.

(which goes beyond mere universalizability and clearly is a substantial moral principle) strongly suggests the former. Therefore, if Lloyd's interpretation were correct, it would seem that my characterization of extra-moral justification does not provide an appropriate account of Hobbes's theory.

While a detailed discussion of Lloyd's book is beyond the scope of this chapter, I think there are strong reasons to stick to the more common interpretation of Hobbes, even in the face of the evidence Lloyd provides in support of her interpretation.³³ Though there are several passages in Hobbes's works about equality, about the need to acknowledge others as equal, and even about some kind of reciprocity, we find him, in virtually all of these cases, offering a prudential rationale for why individuals should allow others equal treatment, a rationale that crucially appeals to the self-preservation of the agent.

Thus, as we have seen, the explanation given by Hobbes for why individuals in the state of nature should conceive of each other as equal is simply that they are sufficiently similar in their physical and psychological properties to serve as a threat to one's safety. Similarly, Hobbes's justification of the laws of nature requiring equity and prohibiting arrogance and pride appeals to the strategic idea that without acknowledging others as equal, other individuals would not be willing to enter into the kind of cooperative enterprise that is needed for securing one's long-term survival. In *The Elements*, Hobbes explicitly emphasizes that nature has ordained the law 'for peace sake ... *That every man acknowledge other for his equal.*'³⁴ In *De Cive* and *Leviathan*, Hobbes makes it even clearer that the requirement follows from the possible threat others pose to the self-preservation of the agent, by explicitly linking it to his previous discussion of the physical and psychological equality of human beings,³⁵ by pointing out that violation will result in a contest for dominion,³⁶ and by emphasizing that an unwillingness to acknowledge others as equal will deprive these others of their willingness to enter into a peaceful state in the first place.³⁷

Moreover, in his references to the Golden Rule, Hobbes avoids any appeal to substantial moral reasons. The Golden Rule is introduced only as a rule of thumb by which individuals may identify the morally advisable course of action without having to enter into

³³ See Lloyd, *Morality*, p. 14ff.

³⁴ EL, 17.1.

³⁵ DCv, 3.13; L, 15, p. 234.

³⁶ DCv, 3.13.

³⁷ DCv, 3.14; L, 15, p. 234.

the sophisticated argument on which Hobbes's derivation of the laws of nature relies.³⁸ The Golden Rule, then, is not presented as an independent foundation or even a negative constraint on the laws of nature but merely as a practical guide for finding out what the laws of nature, and the prudential rationale behind these laws, require in a given situation.

In accordance with this, Hobbes does not seem to have any genuinely moral concept of human dignity or moral worth that could ground a non-prudential moral obligation to treat others as free and equal persons. As Hobbes infamously claims in chapter ten of *Leviathan*, 'The *Value* or WORTH of a man is [...] his Price, that is to say, so much as would be given for the use of his Power: and therefore is not absolute; but a thing dependent on the need and judgement of another.'³⁹ While we may certainly give this definition a somewhat broader interpretation, it would be quite a stretch to attribute to Hobbes anything like Rawls's decidedly moral conception of persons as free and equal. Furthermore, if an asymmetrical solution to the predicament of the state of nature that leaves some individuals with their right to all things should be unacceptable to Hobbes in virtue of the reciprocity theorem,⁴⁰ why should he specifically examine the question of whether the submission of others or the formation of defensive alliances could provide a remedy for the problem of the state of nature and reject these possibilities on purely pragmatic grounds?⁴¹

However, the main problem with Lloyd's interpretation is not that there are no passages in Hobbes's works that allow for it or, if considered in isolation, even suggest it. There is wide agreement among commentators that Hobbes is not always consistent and that we can find isolated evidence for almost any interpretation, especially if we are willing to stray from the main road of Hobbes's argument. The problem is rather to then give a satisfactory account of all the opposed statements and to make sense of the overall structure of his argument. If we take Hobbes's moral theory to depend on the desire for self-preservation (or on self-interest more generally), there are two main explanations we can give for those passages that do not sit well with this interpretation. One is that Hobbes introduced traditional moral vocabulary in order for his theory to be acceptable to a wider audience, the other is that he did so in order for himself to escape prosecution. To some commentators, such explanations may appear *ad hoc*. However, as I have demonstrated above, it cannot be denied that Hobbes appeals to a universal consensus, and given some passages in Hobbes's

³⁸ EL, 17.9f; DCv, 3.26f; L, 15, p. 240ff.

³⁹ L, 10, p. 134.

⁴⁰ For this claim see Lloyd, *Morality*, p. 25.

⁴¹ See EL, 19.3; DCv: 2.5; and L, 17, p. 256.

autobiographical writings,⁴² there can hardly be any doubt that he worried about his safety, either. Moreover, the idea that Hobbes was not therefore willing to give up on *all* revisionary strands in his theory, which he might rightfully have perceived as comprising much of what was original about his theory, is clearly one we need to take seriously. In contrast with this, the commentators who want to turn Hobbes into a more traditional moralist do not seem to provide an equally plausible explanation for the role of self-preservation in Hobbes and for quite a few revisionary statements which make Hobbes appear to be a radical, a pessimist, and perhaps even an atheist when it comes to morality.

What remains to be asked is whether Hobbes is well advised to pursue his justificatory strategy, given his obvious interest in basing his political theory on a consensualist moral theory that allows readers with conflicting religious views to converge on the bare necessities of political life, and the necessity of civil obedience in particular. The *prima facie* counter-intuitive answer for which I argue in the final section of this paper is that it is because of his fundamental approach to justification and his misinterpretations of human nature that Hobbes spoils the potential benefits of his appeal to common sense.

IV.

It is natural to think that when it comes to grounding a moral consensus that can further agreement between groups with different religious or ideological backgrounds, a fundamental extra-moral justification of the kind provided by Hobbes should yield the best results. As long as the descriptive assumptions from which the justification proceeds are as uncontroversial as the human interest in survival (and not as controversial as, for example, the belief in genuine altruistic human motivation), and as long as the ultimate moral conclusions are convincingly derived from these assumptions, it seems that any form of intra-moral justification must stand at a disadvantage—because it takes certain moral principles for granted and we must expect the religious or ideological differences between different groups of readers to include distinct moral convictions.

However, in what follows I want to argue that, if something like this is the idea behind Hobbes's moral theory, he has importantly misjudged both human nature and the demands of moral justification in the political arena. The main advantage which a Hobbesian approach to moral justification can claim over examples of intra-moral justification is that

⁴² See EW, 4, p. 414; OL, 1, p. xv.

such an approach, if otherwise successful, can get hold of the radical moral sceptic who does not yet accept any fundamental moral principles or even as much as a moral standpoint. The figures traditionally cited as proponents of such a radical scepticism include Plato's Thrasymachus, Hume's sensible knave and the Hobbesian 'Foole', and modern examples of Hobbesian contractarianism are explicitly designed as attempts to convince the moral sceptic of the non-moral necessity of moral behaviour.⁴³

Yet, as far as moral justification in the political arena is concerned, the crucial task is rarely, if ever, one of providing an ultimate justification of morality or a refutation of radical moral sceptics. When moral justification becomes a public issue, we are typically dealing with individuals who share some fundamental moral ideas or are at least equally willing to assume a general moral standpoint, since it is this standpoint that gives rise to the need for justification in the first place. To cite a contemporary example, those engaged in modern debates about applied ethics, whose different religious views often bear directly on the issues in question, are not usually worried about the question of whether there are any moral distinctions at all or whether there is an ultimate rational basis for expecting others to act in a particular manner. Rather, they are discussing what the basic moral ideas or values which most of us share and which figure to *some* extent in all major religions or religious moralities (equality, reciprocity, the value of life etc.) can be taken to prescribe with regard to concrete and often unprecedented practices. In a similar manner, we may want to argue that the contentious issue in Hobbes's England was not so much whether civil society is necessary or whether there is any duty of civil obedience whatsoever, since these were points Hobbes's contemporaries widely accepted, their different religious faiths notwithstanding. It was rather to whom such obedience is primarily due and whether it is dependent on any further conditions.

That the ultimate argumentative foundation of moral doctrines is not what decides their public acceptance is also suggested by current research in social psychology. As recent experiments by Jonathan Haidt and his collaborators suggest, 'moral dumbfounding', i.e. clinging to substantial moral convictions in the absence of rational justification, is quite a common phenomenon.⁴⁴ Whatever we may think of Haidt's quite far-reaching anti-rationalist conclusions, the least his findings can be taken to indicate that many human beings are not

⁴³ See Stemmer, *Handeln*, p. 17ff. See also Ann Cudd, 'Contractarianism', *Stanford Encyclopedia of Philosophy*: <http://plato.stanford.edu/entries/contractarianism/>

⁴⁴ Jonathan Haidt, *The Righteous Mind* (London: Penguin, 2013), p. 43.

shaken in their adherence to deeply held moral convictions by their inability to come up with a convincing rationale for those views. If this is correct, however, then it seems that where the idea is to appeal to some kind of universal consensus, a moral doctrine such as Hobbes's will hardly be able to compensate for deviations from the substance of common sense morality by offering a more fundamental justification of the substantial moral principles it actually endorses.

Moreover, it also seems to be a common experience that human beings tend to evaluate moral positions not so much in terms of whether the explicit moral claims associated with this position are in some way consistent with their own convictions, but rather in terms of whether their convictions and their way of life are positively *affirmed* by the position in question. Human beings frequently display hostility towards moral principles that are purely permissive and merely *allow* certain practices, sexual or otherwise, without in any way demanding them. What this suggests is that the property of being less demanding and more tolerant, which we may ascribe to the kind of minimum morality constituted by a Hobbesian approach, does not necessarily add to the social acceptability of a moral doctrine. In fact, the above observation seems also true with regard to those anthropological assumptions on which Hobbes's moral theory is founded. In order for people to accept a proposed view of human nature, it will not be sufficient to define human nature in terms of qualities that are generally included among the qualities of human beings, such as the desire for self-preservation; rather, it will be necessary to include *all* (or almost all) the qualities that others take to be definitive of human nature and to thereby positively affirm their view of what being human is all about.

All this suggests that the fate of the *consensus omnium* envisaged by Hobbes very much hinges on how far his ultimate conclusions positively affirm common sense views on human nature and morality. However, his partial conformity with common sense morality notwithstanding, Hobbes's moral theory is clearly in contradiction with some of our most fundamental intuitions or considered judgements. One important difficulty results from the fact that the moral obligation imposed by the laws of nature is a prudential obligation, whereas we ordinarily conceive of being imprudent and being immoral as two different things. According to common sense morality, the immorality of immoral behaviour is typically linked to disadvantages suffered by others and somehow grounded in their persons and the legitimate claims they have on us. The Hobbesian approach, however, suggests that what is problematic about behaving immorally has not to do, or at least not primarily, with

the interests, desires, rights or even the dignity of others, but with the agent and his own advantage.⁴⁵

A second but related difficulty consists in the fact that the scope of Hobbesian morality seems to be seriously limited, a fact that is explicitly conceded by modern Hobbesians,⁴⁶ and, at least implicitly, by Hobbes as well. Even if we accept Hobbes's claim of relevant equality as far as is needed for a basic defence of the state of nature argument, the rationale behind the laws of nature and their obligatory force does not seem to extend to important groups of human beings which, despite our possibly diverse religious and ideological backgrounds, we all ordinarily take to be members of the moral community. For example, Hobbes's argument fails to securely ground any moral obligations against children or disabled people, a consequence that is implicitly acknowledged by Hobbes himself.⁴⁷ These deviations from ordinary morality derive from the kind of non-moral justification on which Hobbes relies, which does not award human beings any fundamental moral worth or moral dignity from the start, and they are therefore to be found in a similar way in the theories of Gauthier and Stemmer.⁴⁸ If what was said above is correct, however, such deviations pose serious problems when it comes to appealing to a universal consensus, problems that are not outweighed by the possible advantages of Hobbes's approach.

We can sum up these considerations and return to the aspect of religious diversity by appealing to the Golden Rule which is explicitly acknowledged by Hobbes as a valid principle. As is impressively demonstrated by the more recent literature, the Golden Rule is part of the teachings of all major religions. Not only do we find ideas resembling the proverbial dictum *Quod tibi fieri non vis, alteri ne feceris* (on which Hobbes mainly relies) in the Old and New Testaments.⁴⁹ We also find them in the Talmud,⁵⁰ in the works of influential Christian theologians such as Augustine, Anselm of Canterbury, Abelard, Duns Scotus,

⁴⁵ See also Nicholas Southwood who develops his 'normativity objection' against Hobbesian approaches to morality along similar lines, *Contractualism and the Foundations of Morality* (Oxford: Oxford University Press, 2010), p. 34ff.

⁴⁶ See, for instance, Gauthier, *Morals*, p. 268; Stemmer, *Handeln*, p. 257ff.

⁴⁷ See EL, 4.3, 10.8; DCv, 9.2, 17.12; L, 20, p. 310.

⁴⁸ See Gauthier, *Morals*, p. 268; Stemmer, *Handeln*, p. 261.

⁴⁹ Cf. Mt 7.12; and Lk. 6.31.

⁵⁰ Jacob Neusner, 'The Golden Rule in Classical Judaism', in *The Golden Rule. The Ethics of Reciprocity in World Religions*, ed. Jacob Neusner and Bruce D. Chilton (London: Continuum, 2008), p. 59. See also Jeffrey Wattles, *The Golden Rule* (New York and Oxford: Oxford University Press, 1996), p. 42ff.

Thomas Aquinas or Martin Luther,⁵¹ in the Qur'an, in the *hadith* of Muhammad and in Islamic mysticism,⁵² as well as in Buddhism,⁵³ Hinduism,⁵⁴ Confucianism,⁵⁵ and Zoroastrianism.⁵⁶

A fundamental idea of equality or reciprocity, then, is recognized by both the various Christian sects which contributed to religious strife in Hobbes's lifetime and by those religions to which any modern Hobbesian needs to appeal, even though the exact role and import of this element strongly differ—from one religion or religious sect to another as well as from one era in the history of a religion to another. As I have argued above, the fact that Hobbes explicitly includes the Golden Rule in his discussion of the laws of nature can be seen as an expression of his consensualist strategy. Yet, as I have also emphasized, what Hobbes attempts, very much in accordance with his extra-moral approach to justification, is to provide a prudential rationale for the Golden Rule and to thus justify it on non-moral grounds. If the assumptions I have defended in this section are anywhere near the truth, then it should be extremely doubtful whether this brings Hobbes closer to the objective of establishing a moral consensus than if he had simply presupposed the validity of these fundamental moral ideas and worked himself from this presupposition to a more demanding morality with a wider scope. The reactions to Hobbes's theory—and to modern examples of Hobbesian moral theory—strongly suggest that a prudential justification of the Golden Rule will rather repel those who acknowledge the validity and religious dignity of this fundamental ethical principle, especially since this prudential justification has the implication of excluding children and disabled people from the range of the Golden Rule altogether, beings, that is, who, from a religious perspective, are the workmanship of God in just the way healthy mature people are.

The upshot of this is that Hobbes's striving for a moral consensus tends to be ultimately at odds with his idea of establishing a *scientia moralis*. It is plausible to think that Hobbes's reliance on a variant of non-moral justification is not only fuelled by his strive for a

⁵¹ See Wattles, *Golden Rule*, p. 70ff.

⁵² See Th. Emil Homerin, 'The Golden Rule in Islam', in *Golden Rule*, pp. 102, 104f.

⁵³ See Kristin Scheible, 'The Formulation and Significance of the Golden Rule in Buddhism', in *Golden Rule*, pp. 117, 119. See also Charles Hallisey, 'The Golden Rule in Buddhism', in *Golden Rule*, p. 131.

⁵⁴ See Richard H. Davis, 'A Hindu Golden Rule, in Context', in *Golden Rule*, p. 147.

⁵⁵ See Mark A. Csikszentmihalyi, 'The Golden Rule in Confucianism', in *Golden Rule*, pp. 157–169. See also Wattles, *Golden Rule*, p. 15ff.

⁵⁶ See Mahna Moazami, 'The Golden Rule in Zoroastrianism', in *Golden Rule*, pp. 65–75.

moral consensus but also by his preoccupation with the emerging natural sciences and his enthusiasm about the geometrical method. Hobbes's overall aim is to develop a *scientia moralis* that is well integrated into an overall scientific system and can perhaps even be derived from a materialist *prima philosophia*, and it seems that the most promising way to realize this aim might be to build one's moral theory on certain non-moral anthropological assumptions. Contrary to what Hobbes might have thought, however, his two purposes are not equally furthered by one and the same justificatory strategy: the best scientific theory of morality is not necessarily the one that can most easily claim universal acceptance and gain political influence.

My conclusion, then, which, in the absence of comprehensive empirical research into the issues in question, must obviously remain tentative, is that Hobbes misses the target he himself identifies for his moral theory by relying on an idea of moral justification that is not well suited to the aim of grounding a moral consensus in the face of religious pluralism and by underestimating the alternative potential of intra-moral justification. In contrast to what Hobbes and modern Hobbesians suggest, it is doubtful whether being able to proceed from uncontroversial non-moral assumptions represents a relevant virtue when it comes to dealing with religious and non-religious pluralism and whether Rawls's commitment to some kind of moral starting point constitutes any relevant disadvantage.

This is not to say, of course, that the particular way in which Rawls develops his intra-moral approach to justification and incorporates the fundamental ideas of equality and reciprocity is the best one can do. It is a much-discussed question whether Rawls's argument establishes the kind of overlapping consensus between reasonable comprehensive doctrines that it is designed to establish. Moreover, some of the worries concerning the limited scope of Hobbes's morality have been expressed with regard to Rawls's conception of justice as well, if in an importantly qualified way. By interpreting personhood in terms of the idea of society as a system of fair cooperation, Rawls seems to exclude all beings that are permanently incapable of being fully-cooperating members of society from the sphere of justice. It is important to note, however, that this does not apply to children (since they have the potential to develop a sense of justice and a conception of the good) and that even those beings who *are* excluded from the sphere of justice are not thereby excluded from the sphere of morality altogether: As Rawls himself emphasizes, non-fully-cooperating members of society may

impose genuine duties on us—if not duties of justice, then duties of compassion or humanity.⁵⁷

Moreover, even if Rawls may not be in a position to grant children and certain other beings the exact moral role we ordinarily attribute to them, his possible failure in establishing a moral consensus would not be due to the fact that he relies on intra-moral justification, but on the fact that he appeals to the wrong considered judgements or interprets them in an inappropriate manner. The way to improve on Rawls's account, then, would not be to resort to a more Hobbesian approach, but to come up with a more convincing strategy for intra-moral justification. Whether there is such a strategy, and what it looks like, are questions that very much deserve to be discussed in their own right.

⁵⁷ See Rawls, *Theory*, p. 512.