

# 1 Taillight Illumination

## How Rawlsian Concepts May Improve Understanding of Hobbes's Political Philosophy

---

*S. A. Lloyd*\*

In a short piece from 1993 entitled “Some Remarks About My Teaching,” Rawls asserted “we learn moral and political philosophy, and indeed any other part of philosophy by studying the exemplars – those noted figures who have made cherished attempts – and we try to learn from them, and if we are lucky to find a way to go beyond them” (quoted in *LHPP*, xiv). By now, Rawls himself has become one of the exemplars, and we fully expect that future political philosophies will continue to draw lessons from his. The question I wish to explore is whether Rawls's philosophy can provide backward, or what we might call “taillight,” illumination of past philosophical systems. Analogously to the way that the taillights of an automobile let us see features of objects physically behind, can a philosophical system enable us to appreciate features of systems temporally behind?

On the one hand, it wouldn't be so surprising if Rawls's work shed light on ideas in those prior philosophical systems from which Rawls drew to develop his own ideas – say Kant's moral philosophy or Rousseau's social contract – especially where Rawls's formulation of an idea sharpens an idea that was indeed present but only latent, seminal, or inchoate in the earlier exemplars. What may be surprising is that Rawls's work should illuminate our understanding of the political philosophy of a figure like Thomas Hobbes, which provided little of use to Rawls in his construction of “justice as fairness,” except by way of contrast.<sup>1</sup> It would

\* I thank Paul Weithman for his many helpful comments on a draft of this essay, and for the magnificent conference that occasioned its composition.

<sup>1</sup> For instance, the goal of Hobbes's philosophy is to provide all with some or another sufficient reason to submit to a government possessing undivided and legally unlimited powers, rather than to identify principles of justice appropriate for a modern, liberal, democratic, pluralistic society; the “no agreement point” of Rawls's modeling device is general egoism, whereas Hobbes's is the state of nature; Rawls's original position imposes a veil of ignorance whereas Hobbes's state of nature assumes partial knowledge; Hobbes assumes partial compliance in varying degrees rather than strict compliance, as with Rawls; and whereas Rawls characterizes parties as mutually disinterested, Rawls saw Hobbes as characterizing persons as predominantly self-interested and also concerned

be a mistake to suppose that Hobbes could not have influenced Rawls because Rawls thought Hobbes's philosophy unsophisticated or unimportant. To the contrary, Rawls's teaching lectures on Hobbes begin by declaring

In my own view and that of many others, Hobbes's *Leviathan* is the greatest single work of political thought in the English language ... its scope and its acuteness and interesting vividness of observation, its intricate structure of analysis and principles, and its presentation of what I think is a dreaded way of thinking about society which almost might be true and which is quite a frightening possibility ... can have a very overwhelming and dramatic effect on our thought and feeling. (*LHPP*, 23)

And also, Rawls taught, on our philosophizing. He deemed it useful to think of modern British moral and political philosophy as beginning with Hobbes, and with critical reactions to Hobbes, from orthodox Christian moralists including Cudworth, Clarke, and Butler and from Utilitarians such as Hutcheson, Hume, Smith, and Bentham. Rawls said Hobbes's original, formidable, and influential system of thought was for the next century and a half "something in regard to which one *had* to decide where one stood" (*LHPP*, 26). But if Rawls did not develop ideas from Hobbes in his own theory, how could Rawls's theory illuminate Hobbes's philosophy, except again, perhaps by way of contrast?

Part of the answer lies in Rawls's evolving understanding of his own theory of justice between *Theory* and *Political Liberalism* and in his revised understanding of Hobbes's political philosophy from the mid-1980s onward. When *Political Liberalism* came out in 1993, Rawls sent me a copy inscribed "for Sharon, with appreciation and gratitude for your Hobbes, the First political liberal?" (question mark, emphasis on "First"). In his inscription, Rawls was asking me, was Hobbes, supposing your interpretation, the first political liberal?<sup>2</sup> His question is at first puzzling, simply because Hobbes was so clearly not a liberal. Rawls located the historical origins of liberalism in (1) acceptance of the principles of toleration and liberty of conscience, (2) establishment of

with positional standing (*TJ* 1999, 126–27, 211). Rawls was doubtful even that Hobbes's system actually belongs to the social contract tradition Rawls sought to advance (*TJ* 1999, 10, note 4), and I have no doubt that it does not.

<sup>2</sup> Rawls was familiar with my interpretation of Hobbes because he, along with T. M. Scanlon, had supervised the doctoral dissertation in which I began to develop it; and he had read my 1992 book on Hobbes (with evident attention, judging by his copious marginal notes in the copy I had sent him, which Erin Kelly kindly returned to me after his death). Rawls's inscription was dated April 1993; on the 28th of that same month, Rawls wrote in a separate letter "I want to ask you: what might be the predecessors of political liberalism? How far is Hobbes one?"

constitutional regimes of limited monarchy, and (3) support for democracy and majority rule (*LHPP*, 11). Although Hobbes inclines toward (1) if, but only if, the population is willing peacefully to accept religious toleration, he explicitly rejects (2) and sees democracy as a legitimate but nevertheless undesirable form of regime because of its relative instability.

Rawls understood liberal regimes to institutionalize priority for a list of basic rights and liberties, and to guarantee adequate material means to make use of those, in a system that limits executive power and balances it against legislative power. Hobbes, in contrast, criticizes constitutionally limited regimes, especially those with separation of powers, as fragile, prone to paralyzing stalemate and resorting to civil war to overcome it. Hobbes agrees that free exercise of religion would be best if it could “be without contention,” but religious strife was rampant at the time, and Hobbes was not optimistic about the future. Instead, he recommends a church establishment headed by the national civil sovereign, who enjoys ultimate authority to interpret religion’s requirements and to determine external profession and practice. Although Hobbes’s system recommends affording citizens the widest possible scope of liberty compatible with security, equal treatment under the law, and public support of the population, especially the poor, all as duties government owes to subjects under natural law, his absolutism allows for no legal guarantee of these things nor any moral right of rebellion against a regime that fails to provide them. Subjects are released from their duty of obedience only when their sovereign fails to protect them.

As for support for democracy, and rights to vote and to run for office, although Hobbes’s system allows for a democratic form of government, so long as its powers are not limited, it deems democracy positively undesirable because it is more prone than other forms to destabilizing factionalization, to more widely spread corruption and influence peddling (because so many more legislators have favorites to enrich and must be lobbied), and because it elevates demagogues and demagoguery over reasoned political deliberation. This last consideration, that demagoguery eclipses reason when politics must persuade the masses as efficiently as possible, particularly concerned Hobbes. Those are reasons to hope not to live under a democratic regime, although, Hobbes insists, if one already does, it would be both immoral and imprudent to try to replace it with a more stable form of governance. Hobbes then, was no liberal, which, of course, Rawls well knew.

If we instead focus on specifically “*political liberalism*,” Rawls’s question whether my Hobbes was the first political liberal makes much better sense. Political liberalism is characterized by a family of interrelated ideas including reasonable pluralism, a political conception of the person as free and

equal and as possessing capacities for rationality and reasonableness or reciprocity, a freestanding justification of political principles, overlapping consensus, stability for the right reasons, and public reason. If, as I shall argue, viewing Hobbes's system in light of key features of political liberalism – particularly its deployment of a political conception of the person as both rational and reasonable – enables us to appreciate previously unrecognized features of it, Rawls will have provided taillight illumination of a system radically different from justice as fairness.<sup>3</sup> Further, Rawls names and explains some ideas belonging to his political liberalism that we now can see were doing important work in Hobbes's theory on my interpretation of it, lending support to Rawls's suspicion that Hobbes developed a distinctly political justification, in addition to a comprehensive justification, for adhering to his favored principle of political obligation. What might have surprised Rawls though, is that the chief feature of Hobbes's view on Rawls's 1983 interpretation of that view<sup>4</sup> – namely, that it allowed no room for any notion of the reasonable and so no possibility of understanding society as a system of social cooperation rather than of mere coordination (*LHPP*, 87) – turns out to be mistaken. Hobbes's compendious concept of right reason dictates both rational norms of instrumental reasoning and reasonable reciprocity constraints on action. This discovery creates space in Hobbes's system for principled accommodation of a degree of pluralism “in accordance with reason,” compatible with each citizen's committing, for reasons of her own, to a common political principle that enjoys both support in public reason and something like an overlapping consensus of comprehensive doctrines.

### 1.1 Hobbes's Central Problem and Strategy for Its Solution

The central problem Hobbes addresses is political instability, of states collapsing from subjects' withholding cooperation or imploding into civil

<sup>3</sup> Although I believe all six of these ideas are present in Hobbes in at least rudimentary form, I focus here primarily on showing that his conception of the person includes the capacity for reasonableness, both because orthodox interpretations deny this, assuming Hobbesian persons possess only instrumental rationality, and because establishing this capacity is necessary to make room for other aspects of a political liberalism. I briefly indicate how we might understand Hobbes to have offered a freestanding justification for his principle of political obligation, and attempted to demonstrate that it could receive the support of an overlapping consensus, thereby allowing for stability for the right reasons. I discuss Hobbes's conception of public reason elsewhere (Lloyd 2018).

<sup>4</sup> This is the account of Hobbes's theory Sam Freeman transcribed from his tape recordings of Rawls's lectures in Harvard's Philosophy 171 course in the spring term of 1983, supplemented with Rawls's class handouts and handwritten lecture notes, and published in *LHPP*, 23–99.

war as groups compete to harness state authority to promote their own and the society's good as they see it. He was vividly impressed by the "miseries and horrible calamities that accompany a civil war," (*Leviathan* XVIII.20) having lived through one and its chaotic aftermath of rapid regime changes. He attributed that war largely, though not exclusively, to the efforts of diverse religious factions to relocate political power into the hands of those who would use it to institute the requirements of their sectarian religious doctrines. Hobbes comments in his *Six Lessons to the Savilian Professors of the Mathematics* on his writing of *Leviathan*:

the cause of my writing that book was the consideration of what the ministers before and in the beginning of the civil war, by their preaching and writing did contribute thereunto (Hobbes 1839, VII, 335)

and tells us in *Seven Philosophical Problems* that

it was written in a time when the pretense of Christ's kingdom was made use of for the most horrid actions that can be imagined; and it was in just indignation of that, that I desired to see the bottom of that doctrine ... which divers ministers then preached for a pretense to their rebellion." (Hobbes 1839, VII, 5)

Hobbes saw clearly that political stability cannot be reliably secured by force or threat of force. It is not merely that Hobbes deems coercion "all the way down" to be impossible, as he indicates in *Behemoth*:

If men know not their duty, what is there that can force them to obey the laws? An army you will say. But what shall force the army? (Hobbes 1990, 59)

It is primarily that people undertaking political insurrection are often moved by *transcendent interests* – interests that override narrowly self-interested concerns, and for the advancement of which they are willing to die if need be – in causes like religious reformation or in procuring salvation. Those willing even to sacrifice their lives in the service of their larger interest are not likely to be deterred by the sorts of punishments governments have at their disposal. Governments can offer to keep you safe, and can threaten you with a death penalty, but "Eternal life is greater reward, than the life present; and Eternal torment greater punishment than the death of Nature." (*Leviathan* XXXVIII.1)

If the command [of the civil sovereign] be such as cannot be obeyed, without being damned to eternal death; then it were madness to obey it." (*Leviathan* XLIII.1)

In *Philosophical Rudiments (De Cive)* Hobbes had observed: "Neither is any man so mad, as not to choose to yield obedience rather to them who can remit and retain their sins, than to the powerfulest kings" (Hobbes 1839, II.17.25). "For every man, if he be in his wits, will in all things yield that man an absolute obedience, by virtue of whose sentence he

believes himself to be either saved or damned” (Hobbes 1839, II. 18.14). Government sanctions can’t compete with divine sanctions; and even sanctions aside, “it is manifest enough that when a man receiveth two contrary commands, and knows that one of them is God’s, he *ought* to obey that, and not the other, though it be the command even of his lawfull Sovereign” (*Leviathan* XLIII.1, emphasis added).

Political regimes cannot hope to remain stable without buy-in from adherents of the main religious or moral doctrines present in the society, which requires that society’s members see political obedience to the existing regime as, at bare minimum, compatible with those commitments, and better, as positively supported by them. Hobbes’s proposed solution was to identify a principle of political obligation – a principle specifying the conditions under which a citizen is to obey the political authority under which she lives – that could, if widely followed, reliably secure domestic peace and then provide each citizen with what she could regard as a *sufficient* reason to adhere to that principle.<sup>5</sup>

However, different people have different values and interests, reflecting differences in their bodily constitutions, upbringings, experiences, habits, education, and self-conceptions, to such a degree that it is “impossible that ... all men consent in the desire of almost any one and the same object” (*Leviathan* VI.6), and the “*objects* of the passions, which are the things *desired* ... the constitution individual and particular education do so vary... as they are legible only” to God (*Leviathan* Introduction). Given this inevitable pluralism, Hobbes saw that providing each person a reason not overridden by contrary reasons required him to show that his proposed principle of political obligation advanced a *multitude* of standard types of interest, including: (i) prudential interests in safety and commodious living, (ii) moral interests in fulfilling one’s natural duties and voluntary obligations, (iii) the religious interest in fulfilling one’s duties to God, and (iv) “special-prudential” interests in

<sup>5</sup> The principle he aimed to establish was that one should obey an existing sovereign so long as it is effective in protecting one, in all of its commands excepting only those that would require violation of one’s duty to God. As Hobbes summarizes his principle (omitting the effectiveness condition on political obligation) “subjects owe to sovereigns simple obedience, in all things wherein their obedience is not repugnant to the laws of God” (*Leviathan* XXXI.1). In *De Cive*, Hobbes says of the principle “it has been shown, both by natural reason and from holy scripture, that citizens should obey Princes and rulers of the commonwealth in all things, except in what is contrary to God’s commandments” (Hobbes 1998, 18.13). Hobbes carried out his attempt to reconcile religious and special-prudential interests with his principle of political obligation in the second half of *Leviathan*. There he argued that God’s natural law requires obedience to the commands of an effective sovereign and that revealed religion as contained in Scripture makes obedience to the sovereign’s (even erroneous) commands concerning religion one of two necessary conditions for salvation (the other being belief that Jesus is the Christ).

receiving salvation or avoiding damnation. Hobbes's effort to provide a *confluence* of distinct reasons converging in support of his proposed principle of political obligation looks to be an effort to show that there can be something like an overlapping consensus on his principle, allowing it to enjoy the principled commitment of diverse individuals embracing various moral or religious doctrines, thereby making their allegiance to it more stable and the society ordered by it more than a mere *modus vivendi*. This may be one of the features of Hobbes's view on my interpretation of it that set Rawls to wondering whether Hobbes may have been the first political liberal.<sup>6</sup>

I have emphasized Hobbes's insistence on the pluralism of human ends – a fact not often appreciated by standard interpretations – because it limits the sort of “political” conception of the person and of society available to Hobbes. He asserted in an early version of the theory that the desire to avoid bodily death is the strongest of desires “from nature” (meaning, presumably, biologically hardwired) in every (healthy) human being. However, because Hobbes recognizes that this natural desire is often overridden by competing desires both natural and “from culture,” including the desire to escape an unacceptable quality of life, desires for glory, honor or reputation, for divine rewards or to avoid divine punishments, and desires to advance valued ends including the good of the nation and the interests of loved ones, Hobbes does not implausibly

<sup>6</sup> Although in *Political Liberalism* Rawls leaves it to adherents to various comprehensive doctrines to develop a supportive connection between their doctrine and the object of overlapping consensus, in *Theory* he offers an example of a possible such connection in his “Kantian interpretation” of justice as fairness. Of course, that interpretation may be contested by other Kantians, just as we might expect different Utilitarians to disagree as to how best to relate utilitarianism to justice as fairness. Reasonable disagreement is possible. Hobbes offers his own interpretation of Judeo-Christian religion in support of his principle of political obligation, as this is necessary in order to challenge the interpretations of those who would vest authority to interpret religious requirements in someone other than the civil sovereign, e.g., Independents (who privileged the individual's interpretation of Scripture) and Roman Catholics (who privileged the Pope's judgment). He needs to show in detail how his principle of political obligation is both compatible with and supported by Scripture, if he is to reassure his Christian audience that they may defer to their sovereign without compromising their eternal prospects or failing in their duties to God. A. P. Martinich has persuasively argued that Hobbes's own interpretation is fairly typical of an English Calvinist of the period, once Hobbes corrected a couple of eccentric positions driven by misapplications of his philosophical innovations (Martinich 2021). For Hobbes, reasonable disagreement about the requirements of religion is possible because natural religion based on reason alone cannot settle all questions of interpretation of revealed religions' authoritative texts. However, it can settle some questions and rule out some interpretations; thus, not all interpretations of a revealed religion are equally reasonable, and some may be unreasonable. Such a view is consistent with a political liberalism.

stipulate the universality of an overriding desire for self-preservation.<sup>7</sup> Desire for a pleasant, commodious life is common; yet many are prepared to forgo such a life in pursuit of other ends. “All men agree on this, that peace is good” (*Leviathan* IV.40), but peace is often deemed less good than other ends by “needy and hardy men” seeking military advancement or by those who prefer victory to peace – or those who do not want peace on just *any* terms (comprising most of those prepared to fight and many of their civilian supporters). Hobbes does stress the strength and ubiquity of the desire to be *highly esteemed* by others, but insofar as high interpersonal standing is a positional good, this desire seems unpromising as the centerpiece of politically liberal conceptions of citizens and society. What Hobbes needs to ground such conceptions is a desire no person can fail to have, and which is never overridden.

## 1.2 A Political Conception of the Person?

In his teaching lectures, Rawls suggested that Hobbes may not have intended the conception of human nature he articulated to be strictly true but rather as appropriate for the limited purpose of devising a theory of political obligation (*LHPP*, 46). Although acknowledging that individuals are capable of benevolence (and so rejecting psychological egoism), Hobbes thought that for political purposes we ought not to assume that people will sacrifice their interests for the sake of strangers or fellow subjects but should instead assume them to be predominantly self-interested. Rawls read Hobbes as attributing to all men interests in self-preservation, conjugal affections, and commodious living – in that order – and as possessing the instrumental rationality to enable them to appreciate that submission to government could be expected better to advance those interests than could life in a state of nature. Rawls interpreted Hobbes’s laws of nature as articles of a shared “secular morality” meant

<sup>7</sup> Indeed, if Hobbesian men really cared most for their self-preservation, we should not expect to see them engaging in the life-threatening activity of active rebellion – war being so obviously hazardous to one’s health – except perhaps against extraordinarily violent regimes attacking themselves or extraordinarily weak regimes exposing them to violence from others. The fact that Hobbes views seditions and rebellions against powerful coercive states as frequent happenings and an ever-present possibility suggests that he did not see fear of death as the main motivator of action. The prominent role of fear of death in Hobbes’s political philosophy is not to provide a necessary end of action but is rather to explain why we would judge others unreasonable to fault us for trying to defend our life and relying on our own judgment as to how to do so – “it is therefore neither absurd nor reprehensible, neither against the dictates of true reason, for a man to use all his endeavors to preserve and defend his body” (*De Cive* 1.7). Combined with the reciprocity requirement, this judgment yields the universal right of nature. See Lloyd (2009, 63–73).

as a political doctrine (LHPP, 51, 53), thinly justified as instrumentally advancing those interests. He saw the content of those articles as enjoining reasonable constraints on conduct; but Rawls understood Hobbes to have offered a strictly rational justification for the law of nature's reasonable precepts. Because understanding society as a system of social cooperation rather than of mere coordination requires attributing to participants a capacity for mutuality or reciprocity (LHPP, 56, 62), which those lacking the capacity for reasonableness do not have, Hobbes's system could be at most one of social coordination. Rawls concluded "if moral right and obligation involves grounds different from the Rational, as I believe it does, Hobbes has no place for it in his *official* view" (LHPP, 66) and that "Hobbes has no place for a *sense of fairness*" (LHPP, 87).

Let me now offer new support for Rawls's characterization of Hobbes as having advanced a political conception of the person, while arguing for a quite different characterization of the interests and capacities comprising that conception. Hobbes's method of demonstration requires that his premises be either conceptual truths settled by definition or empirical truths confirmable by introspection or observation – premises "as passion not mistrusting, may not seek to displace" (Hobbes 1839, IV, dedicatory epistle). On my account, Hobbes assumes:

- (1) People have ends they desire to advance by their actions.
- (2) People desire that the conditions necessary for their actions to be effective in achieving their desired ends obtain.

These premises are entailed by Hobbes's definition of man as a "rational animal," which definition he believes is widely accepted. *Qua* animals, people have internal appetites and aversions suggesting ends of their voluntary actions; *qua* rational, people seek to secure the conditions necessary for their actions to achieve their ends.<sup>8</sup> Putting this second

<sup>8</sup> Cases in which we might be tempted to say that we desire an end but do not desire that the conditions needed to achieve that end obtain are best understood as cases in which we have not adequately characterized our desired end. When I desire to win the presidency but do not desire the election fraud that would be needed for me to win the presidency, what I actually desire is to win the presidency "fair and square"; and of course, I *would* want the conditions necessary for achieving that end to obtain. If I desire to inherit my parent's wealth, but not the requisite death of my parent, my first-order desire is better characterized as the desire to inherit when the time comes. Hobbes explicitly affirms as a principle of rationality that he who wills the end must will what he believes to be the necessary means to that end; that principle suggests a tight connection between the necessary desire and every first-order desire, so *if* the necessary desire is a second-order desire, it will be a special kind of second-order desire. Unlike ordinary second-order desires – e.g., "I desire that I shall desire to eat healthy foods rather than the fast food I actually desire – which are not entailed by the first-order desire they govern – the necessary desire does seem to be implied by any first-order

point negatively, insofar as a person is rational, she desires to avoid a condition that can be expected to undermine her ability to act effectively. Hobbes's subsequent argument explains why a state of nature is just such a condition. It is a condition in which mutual interference and insecure control of resources threaten achievement of any of our ends, no matter what those ends are. This explains why in each presentation of his state of nature argument, Hobbes adduces a long list of desirable ends unlikely to be realizable – achievements in arts, letters, sciences, geographical knowledge, society, elegance, comfort, safety, riches, freedom from fear, etc.

Hobbes views the desire to achieve one's particular ends, and the desire to secure the conditions required for doing so –which I term “the necessary desire” – to be inescapable for any rational human. Because he intends his state of nature argument to demonstrate to *everyone* that they have at least one very good reason in common to submit to an effective political authority, it is crucial that that argument not depend on premises that are false of some people because it attributes to them desires they lack, or which in them are overridden by other desires. The “necessary desire,” by not presupposing possession of desire for any particular object, is as thin and as unladen with controversial values as Hobbes can make it.<sup>9</sup>

One advantage of this understanding of the foundation of Hobbes's political philosophy over Rawls's 1983 interpretation is that it provides a more clearly political (because less contentious) conception of the person upon which Hobbes can build his argument for political obligation. Rawls saw that Hobbes was “attempting to identify fundamental interests by which everyone is moved” (*LHPP*, 67). Hobbes's actual identification – that we have an interest in securing the conditions necessary for our actions to be effective in achieving whichever ends we actually have, is far less controversial than assuming that in everyone, “our interest [is] first in preserving ourself” (*LHPP*, 46). Hobbes's assumption of a desire

desire I really do have. Thanks to Susanne Sreedhar for encouraging me to think about the status of the necessary desire.

<sup>9</sup> Hobbes indirectly expounds this desire in his discussion of the desire for “powers,” understood as one's present means to obtain some “future apparent good” (*Leviathan* X.1). People with desires must want powers (among which Hobbes includes good luck in external circumstances and cooperative allies such as friends, servants, and admirers, who help to advance one's ends). Hobbes's claim that people seek “power after power” (*Leviathan* XI.2) in order to “assure forever” the satisfaction of their future desires is best understood as the quite plausible claim that people sequentially seek various powers over time as they come to foresee the ends they are likely to have, rather than as attributing to each an effort to gain maximum power at every moment. An orderly social environment, by reducing the impediments posed by other people, reduces the types and magnitudes of powers needed to achieve one's ends.

to act effectively is also less contested than an assumption of merely predominant egoism; and the fact that it is analytically derivable from our nature as a rational animal, (in contrast to the primacy of self-preservation), conforms to Hobbes's stated "scientific" method.<sup>10</sup>

What Hobbes terms "the condition of mere nature" is a condition in which it is *morally permissible* for each individual to act on their own private judgment in every matter. In characterizing his imagined pre-political condition this way, Hobbes alludes to two further features of his conception of the person for the purpose of establishing political obligation: persons are free and equal. Were persons born under an obligation to defer to the judgment of others, including to political authorities, action on their own divergent private judgments would be impermissible. Although Hobbes personally believes that children in fact owe obligations of obedience to the person who preserved them alive – always in the first instance to their mother – for political purposes, people are to be understood to be each and equally at liberty to make and to act on their own decisions. In the condition of mere nature, there is a universal right of each to act on private judgment in every matter.

Hobbes does not offer as the basis of equal liberty some theory-laden claim about humans' possession of innate dignity, or status *qua* God's children. Instead, he calls our attention to the observable fact that as adult individuals, we are sufficiently similar in the sorts of capacities needed to impose our will on others or to resist imposition of theirs on us – bodily strength, intelligence, ingenuity, wiles – none of us can reasonably expect to get our way against everyone else. Even if any one of us were marginally superior to others in *all* the qualities mentioned, others can *combine* their talents to outdo us. People who are operationally equal will have similar hope to achieve their ends and will see no reason why they should defer to others' judgments, rather than act on their own. Hobbes writes:

In the faculties of body and mind ... the difference between man and man is not so considerable as that one man can thereupon claim to himself any benefit to which another may not pretend as well as he. (*Leviathan* XIII.1)

If nature therefore have made men equal, that equality is to be acknowledged; or if nature have made men unequal, yet because men that think themselves equal will not enter into conditions of peace but on equal terms, such equality must be admitted. (*Leviathan* XV.21)

This basis of equality is thinly political. The reason for assuming natural freedom is similarly thin: natural equals have no cause to accept

<sup>10</sup> If Rawls's idea of a political conception also requires that it be worked up from the public political culture, then Hobbes embraced a more minimal condition.

purportedly natural normative relations of domination and subordination. Hobbes argues that if it is permissible for you to judge all matters, including matters pertaining to me – such as the correctness of my judgments about how to act – it must be equally permissible for me to judge matters pertaining to you – including the correctness of your judgment about the correctness of my judgment as to how to act.<sup>11</sup>

The final element of Hobbes's conception of the person is, like the necessary desire, entailed by his expansive notion of reason: *qua* reasoning creatures, humans adduce reasons, or justifications, for their actions (and beliefs), which reasons are subject to a *consistency constraint*. Not only does reason forbid simultaneously affirming contradictory beliefs, it also prohibits affirming and acting on contradictory practical principles. The ability to hold oneself to the same standard of practical action one applies to others – the capacity for *reciprocity* – is a central feature of Hobbes's conception of the person; this ability makes it possible for Hobbesian persons to comply with the most basic natural law requirement – reciprocity – the common core from which all other laws of nature are derived. Hobbes refers to the reciprocity requirement variously as the “sum” of the laws of nature, as “containing” all the laws of nature, or says that it “just is” the law of nature. He formulates it more than a dozen times, both positively and negatively as dictating

Whatsoever you require that others should do to you, that do ye to them. (14.5)

Do not that to another which thou thinkest unreasonable to be done by another to thyself. (*Leviathan* XXVI.1)

In its political application, reciprocity demands

That no man require to reserve to himself any right which he is not content should be reserved to all the rest. (*Leviathan* XV.22)

Notice that this is a “weak” reciprocity requirement in the sense that it never requires you to adopt the behavioral standards of *others*; someone else's choice to treat you according to some general principle *they* propose does not impose any duty on you to conform to that principle in your treatment of them. Nor does it require you to adopt universalizable standards that could be adopted by all from some impersonal point of view. Rather, your own demands or evaluative judgments or practical attitudes set a standard for action that reason then requires you to apply

<sup>11</sup> Hobbes writes, “say that another man is judge. Why now, because he judgeth of what concerns me, by the same reason, because we are equal by nature, will I judge also of things which do belong to him. Therefore it agrees with right reason, that is, it is the right of nature that I judge of his opinion.” (*De Cive* I.9)

across the board – including to your own actions – and to accept from others as justifying their like actions, even in circumstances when doing so does not advance your self-interest.

I mentioned earlier that Rawls viewed Hobbes's laws of nature as a "secular morality" offered as part of a political conception. Some further evidence for that view, which Rawls does not mention, is that the moral virtues expressed in the laws of nature do not include any traditional moral virtues that redound to the personal benefit of the agent who possesses them without necessarily benefiting her political community. Hobbes excludes temperance, prudence, and courage from among the dispositions required by the laws of nature precisely because those virtues *in an enemy* may prove detrimental to the flourishing of our community. These "are not virtues of citizens as citizens, but as men," whereas "good manners (that is, moral virtues)" are dispositions that facilitate and preserve civil societies (*De Homine* XIII.9). Hobbes's laws of nature pick out only virtues or vices of persons engaged in social relations<sup>12</sup> and not of humans per se.

Hobbes's reciprocity requirement is not itself justified instrumentally as necessary for securing self-interest and is thus not reducible to "the rational." In fact, violating the reciprocity requirement is not always irrational, because, as Hobbes concedes, some people, in particular powerful people, can correctly expect to profit by doing so; Hobbes acknowledges that in this life (the only one we can observe) the wicked do prosper, many a time. Reciprocity is an "eternal and immutable" requirement binding at all times on everyone with the use of reason: Yet if its justification depended on its unfailingly serving self-interest, it could not have that status.

Compare Hobbesian reciprocity with Rawls's standard of reasonableness – which is that we be ready to propose and to abide by terms of cooperation we think others could, as free and equal persons, reasonably accept, so long as others also honor those terms. Hobbesian reciprocity *does* require willingness to abide by the standards we propose, and, because we believe ourselves (who are proposing them) to be reasonable, and we accept those standards, we obviously think a reasonable person could accept them. Rawls's worry is that differences in bargaining position may enter here, and it is true that Hobbes makes no provision to abstract from those; but recall that Rawls was devising a theory of justice, whereas Hobbes was pursuing the different project of justifying submission to an absolutist government. If we agree that even imperfectly just

<sup>12</sup> The law of nature requiring allowing safe passage to mediators would apply to relations among political communities and between hostile factions within a political community.

governments may still be better for securing human interests than no government at all, we should not define the reasonableness requirement for social cooperation so narrowly that no arrangements short of perfectly just ones will count as any form of social cooperation. Hobbesian reciprocity enables principled, internally motivated cooperation, as opposed to mere coordination imposed by external coercion.

Reason's reciprocity requirement has a deep hook in human psychology as Hobbes understands our psychology: Humans care very much to be and to be seen by others as being *justified* by reason in their actions, as each can introspectively confirm.<sup>13</sup> Hobbes's persistent insistence on the influence on our social interactions of our pride, self-conception, and concern for status (which he discusses under the headings of glory and vain glory as well as pride) explains why we feel it degrading to be exposed as deviating from a requirement of reason. Our higher ability to reason is what distinguishes adult humans from nonhuman animals. Actions "that proceed from error, ignorance, or folly [are] dishonorable" (*Leviathan* X.42); and "craft, shifting [cheating], neglect of equity is dishonorable" (*Leviathan* X.46). We take offense when others disagree with us on multiple matters because we see such disagreement as impugning our status as intelligent reasoners. Hobbes makes clear that the main way in which people depart from reason's requirements is by hypocritically holding others to different practical standards of conduct than they apply to themselves.

Rawls rightly observed that Hobbes's laws of nature, both the reciprocity requirement itself and the specific secondary laws dependent on it for their derivation, "define a family of *reasonable* principles so far as their content and role discern" (LHPP, 64). But we can now see that reciprocity "accords with reason," not because, or only when, it instrumentally serves personal desires, but rather because reason imposes a *consistency constraint* on our judgments and attitudes as well as on our beliefs. Reason dictates in more than one way – forbidding both logical contradiction and failure to fit perceived means to ends. Hobbes criticizes the reputedly wise Cato on the ground that with him "animosity should so prevail instead of judgment, and *partiality instead of reason*, that *the very same thing* which he thought just in his popular state, he should censure as unjust in a monarchical." And the fact that one "gives a different judgment of an action when he does it than when someone else does the very same thing ... [is among] the obvious signs that what moral

<sup>13</sup> For discussion, and an argument that Hobbes views the desire for self-justification as a potentially powerful motive for conformity with the moral norms articulated in his laws of nature, see Lloyd (2020).

Philosophers have written up to now has contributed nothing to the knowledge of truth" (*De Cive* Dedicatory Epistle).

It is contrary to reason to hold contradictory attitudes toward the same justifying consideration, or reason, for a fixed action-type (an action under a general description). When we are acting, not on instincts or whims, but on proffered reasons, we are committed to the consistency requirement articulated by the reciprocity rule. Hobbes illustrates violation of the reciprocity requirement with the biblical story of the prophet Nathan's rebuke of King David over his appropriation of Uriah's wife, Bathsheba. Nathan poses David's own action to David by way of an analogous case to be judged of a rich man who had many lambs but chose to sacrifice the only lamb of a poor man: When David judges "the man that hath done this *thing* shall surely die" the prophet answers, "Thou *art* the man." David is rebuked – on Hobbes's reading – not for coveting another man's wife, nor even for causing Uriah's death in order to gain her, but rather for *inconsistently* approving in himself *the very same type* of action he condemns in another (the rich man of the story).

Another striking example is Hobbes's application of the reciprocity requirement to defend his position that whatever a subject like the Christian convert Naaman, who bowed before an idol, does, not because he approves it but because the civil law requires it, is the action of his sovereign, and not of himself, and so is permissible. Hobbes reasons

I ask [any objector], in case there should be a subject in any Christian Commonwealth that should be inwardly in his heart of the Mahomedan Religion, whether if his sovereign command him to be present at the divine service of the Christian Church, and that on pain of death, he think that Mahomedan obliged in conscience to suffer death for that cause, rather than to obey the command .... If he say he ought rather to suffer death, then he authorizeth all private men to disobey their Princes, in maintenance of their religion, true or false; if he say, he ought to be obedient, then *he alloweth to himself, that which he denyeth to another*, contrary to the words of our Saviour "Whatsoever you would that men should do unto you, that do ye unto them," and *contrary to the Law of Nature*, (which is the indubitable and everlasting Law of God) "Do not to another that which thou wouldest not he should do unto thee." (*Leviathan* XLII.11), emphasis added)

More generally, Hobbes appeals to the reciprocity requirement in his condemnation of proselytizing efforts to convert foreign populations to our religion: The missionary "does that which he would not approve in another, namely, that coming from hence, he should endeavor to alter the religion there" (*Leviathan* XXVII.4).

Notice that none of these condemnations of violation of the reciprocity requirement appeals to instrumental irrationality. The criticism is not

that one is failing to take necessary means to one's own ends. The criticism is of inconsistency in one's practical principles – a type of hypocrisy – in affirming moral standards by applying them to others, while not conforming one's own behavior to those standards (not merely as a matter of weakness of will), but denying their application to oneself. What goes on in these violations of reason's reciprocity requirement?

Hobbes characterizes reasoning as calculative, as adding or subtracting “names in our affirmations,” in a syllogistic process.<sup>14</sup> We can make sense of Hobbes's condemnation of practical inconsistency if we understand him as holding that when a person offers a reason or justification for some action, evaluative attitude, or practical judgment – let's call this an “item” – she is committing herself to a *general* claim – no indexicals or definite descriptions allowed – from which her item is supposed to follow by syllogistic argument. For example, taking “because” to mark a proffered reason or justification

(a) I fault your behavior **because** *it is harmful*

commits me to the general claim.

(A) *harmful behavior is faultworthy.*

By Hobbes's account, if the “because” clause is to provide a candidate reason, it must follow deductively from a general claim/principle that properly interrelates its component terms. Thus, only if “is faultworthy” contains everything “named” by “is harmful” has (a) expressed a reason for faulting your behavior. This creates a tight link between Hobbes's conceptions of reasoning and of having/being/offering a reason.

However, this link operates *in both directions* between general principle and specific conclusions, so the general principle that justifies my particular “item” will also justify anyone else's like item. And my tacit appeal to that principle in claiming reason for my own item commits me to acknowledging equal reason for anyone else's like item. So, if I offer (a), I then behave contrary to reason when I refuse to fault my own harmful behavior or refuse to accept from you that the fact that my behavior is harmful is also a reason for you to fault it. This is because those practical stances, when converted into propositional form – (a') my harmful behavior is not faultworthy, (a'') that my behavior is harmful is no reason for you to fault it – directly contradict the general claim (A) upon which my proffered reason depended.

All right reasoning in any domain depends on consistency. The reciprocity rule expresses reason's requirement of consistency in the domain of

<sup>14</sup> He writes “that making of syllogisms is that we call RATIOCINATION or reasoning.” (*Elements* I.5.11)

the practical. That is the claim it makes on humans, for whom, *qua* “rational animals”, reason is normative. The reciprocity requirement is not justified instrumentally, as contributory to the satisfaction of any desire, and so is not reducible to a tenet of means-ends rationality. It expresses a requirement of reasonableness, which Hobbesian persons do accept and which they have the capacity to meet, completing Hobbes’s conception of the person as free and equal, rational and reasonable, and having the desire to secure the conditions necessary for acting effectively in pursuit of the ends she has. This does look like an appropriately political conception of the person.

### 1.3 A Better Argument from the State of Nature for Submission to Government

Assembling these elements, we can now state Hobbes’s argument: In the state of nature, (1) each individual has moral liberty to act on her own private judgment in every matter, with no duty to defer to anyone else. In this sense, persons are free. (2) Persons are also equal, meaning that none can rationally expect to impose his will over others or “claim any benefit they cannot claim as well.” (3) Persons have ends they desire to achieve by their actions, and, as rational, (4) they desire that conditions for their actions to be effective in achieving their ends obtain. (5) Persons are capable of reciprocity – of holding themselves to the same standards they demand others meet – and insofar as they act in conformity with reason, abide by the reciprocity requirement.

Hobbes then argues: Because individuals not infrequently *disagree* in their practical judgments and pursue incompatible ends, (as experience confirms), the condition of *universal* private judgment can be expected to result in mutual interference, and insecure access to needed resources (including the cooperative aid of others), that seriously compromises any given person’s prospects for acting effectively to achieve their ends. Rational persons *must* therefore demand that *others* give up their right to act according to their own private judgment in everything. But, by reciprocity, what we demand of others we too must do. Thus, reason dictates “that a man be willing, when others are too, to give up his right to all things, as far as he thinks necessary, and be contented with so much liberty against other men as he would allow other men against himself” (*Leviathan* XIV.5). This is Hobbes’s second law of nature, operationalized by undertaking political obligation, because the only reciprocal alternative to governance by universal *private* judgment is universal deference in some range of matters to a *public* judgment that can arbitrate disputes and enforce rules to create a navigable social environment in which we can expect to effect our ends.

This certainly looks like an argument “freestanding” from controversial comprehensive doctrines.

## 1.4 Conclusion

Rawls’s query whether Hobbes may have been the first political liberal is tough to answer partly because it runs so contrary to the understanding of the basic contours of Hobbes’s political philosophy that has been dominant for a good three hundred years and reinforced in the latter half of the twentieth century. There was an influential school of interpretation, beginning now fifty years ago with David Gauthier in the late 1960s and developed by Greg Kavka and Jean Hampton in the 1980s, that applied a thick layer of alien rational choice theory and game theory over Hobbes’s original argument that assumed Hobbesian persons to be narrowly rational, wholly or predominantly self-interested actors.<sup>15</sup> My argument here has indicated how those interpretations distort Hobbes’s political philosophy beyond recognition. They are false to the complex psychology of Hobbesian persons (and of us); they completely obscure the moral basis of Hobbes’s system; they oversimplify his conception of the operations of reason; and they would render ineffective Hobbes’s primary recommendation for preserving a well-designed commonwealth, which is, not to change the payoff matrix by instilling terror in the population, but to reform university education and pulpit preaching to acculturate citizens in their moral, political, and religious duties (*Leviathan* XXX.4, 7–14; Hobbes [1990] [*Behemoth*] 39–40, 56, 58–59).

I was reminded not too long ago of misunderstood Hobbes by a *New York Times* report that Johannes Vermeer’s famous painting, “Girl Reading a Letter at an Open Window” turns out to contain a naked Cupid hidden under a rectangle of paint that makes the wall to the letter-reader’s right appear totally bare, empty.<sup>16</sup> This fact had been discovered forty years ago by X-ray analysis of the painting, but it was only recently, when museum restorers applied a solvent to the paint rectangle, that they realized it must have been applied later by a foreign hand and that the

<sup>15</sup> Even Rawls, in his 1983 teaching lectures on Hobbes, described Hobbes’s state of nature as having a structure analogous to the prisoner’s dilemma game (*LHPP*, 73–78, 88–90), which explained why the advantages of adhering to agreements and observing the requirements of natural law are unavailable to individuals living without a sovereign enforcer. Less guardedly, Rawls (1999, 238), states that “Hobbes’s state of nature is the classical example” of “the general case of the prisoner’s dilemma.” See also Gauthier (1969, 79–85); Kavka (1986, 109–13) and *passim* and earlier articles; Hampton (1986); Curley (1994, XXIV–XXV).

<sup>16</sup> [www.nytimes.com/2021/09/09/arts/design/vermeer-cupid-restoration.html?smid=url-share](https://www.nytimes.com/2021/09/09/arts/design/vermeer-cupid-restoration.html?smid=url-share).

Cupid was in fact Vermeer's, a realization that altered and enhanced scholars' understanding of the painting's intended meaning. (For instance, that the girl was more likely reading a love letter than a vender's demand for payment).

I am hopeful that a similar sort of lost-and-found is happening in Hobbes studies. However well-intentioned the application of the shiny new tools of game theory to Hobbes's political philosophy may have been, the result was to obscure some of its most important elements, including its appeal to citizens' reasonableness as a central and irreducible presupposition of the theory and its recognition that citizens' transcendent religious or moral interests can be significant sources of social disorder.<sup>17</sup> Thinking now in the Rawlsian terms of political liberalism about Hobbes is not *exactly* like applying solvents to paint overlays on an Old Master, but it does aid us in washing away the innovative additions of Hobbes's later interpreters. Rawls's political liberalism provides "tail-light illumination" of Hobbes by enabling us to notice those thick and controversial assumptions about human nature and motivation both Hobbes's admirers and detractors coated him in and by supplying tools for understanding what Hobbes was doing that reveal him to have been more original and interesting than we had imagined.

<sup>17</sup> One might wonder whether such prisoner's dilemma arguments could helpfully supplement Hobbes's system, not in their advocates' intended role of providing the state of nature argument for submission to government but by assuring us that reasonability and rationality align for the most part, so that we can see our commitment to reciprocity as a part of our good, in turn enhancing social stability by limiting defections from cooperative norms. When Rawls addressed the question in *A Theory of Justice* whether having a sense of justice is a part of our good according to the thin theory, he asked this about members of a well-ordered society who have *already* acquired a sense of justice. Hobbes, in his famous reply to the Foole, engages the more ambitious project of showing that even an atheist without a sense of justice who rebels unjustly acts "against reason" and in discordance "with his own good" (*Leviathan* XV.5, 91 note 5, Latin variant) because she relies on a faulty rule of inference (viz., if an action turns out well, it cannot have been against reason to perform it), and because she incorrectly extrapolates from experience (expecting to go undetected even though most such deceptions are discovered). This reply depends *in no way* on game theory modeling of strategic choice. The Foole simply lacks both sapience and prudence and is, as such, a defective person. See Lloyd (2009, chapter 7). In contrast, ordinary people accept the natural law requirement of reciprocity as a dictate of reason and would be embarrassed, ashamed, or offended to seen by others as incapable of reasoning, or as too weak of will to act on their reason, or as hypocrites who endorse for others requirements they are unwilling to accept themselves. For Hobbes, part of our good is to be, and to be acknowledged as being, not overall inferior to others; this provides a natural motive for trying to live up to reason's requirements. For discussion of how the desires for self-admiration and the respect of others may motivate compliance with the moral requirements articulated in the laws of nature, see Lloyd (2020).