

# Hobbes's Struggle with Contractual Obligation. On the Status of the Laws of Nature in Hobbes's Work

Matthias Kiesselbach

Institut für Philosophie, Humboldt-Universität zu Berlin, Unter den Linden 6,  
10099 Berlin, Germany

E-mail: Matthias.Kiesselbach@philosophie.hu-berlin.de

**Abstract:** This paper argues that throughout his intellectual career, Hobbes remains unsatisfied with his own attempts at proving the invariant advisability of contract-keeping. Not only does he see himself forced to abandon his early idea that contractual obligation is a matter of physical laws. He also develops and retains doubts concerning its theoretical successor, the doctrine that the obligatoriness characteristic of contracts is the interest in self-preservation in alliance with instrumental reason – i.e. prudence. In fact, it is during his work on *Leviathan* that Hobbes notes the doctrine's main shortcoming, namely the limitation of its dialectical potential to cases in which contract-breakers are publicly identifiable. This essay shows Hobbes's doubts about his *Leviathan's* treatment of contractual obligation by way of a close reading of its central 15<sup>th</sup> chapter and an analysis of some revealing shifts between the English *Leviathan* and the (later) Latin edition. The paper ends by suggesting that Hobbes's awareness of the flaws at the heart of his political philosophy helps account for some striking changes in his latest writings.

**Key Words:** Contract-Keeping, Obligation, Laws of Nature, Rationality, Hobbes, Leviathan

**This is a penultimate draft! Please cite from the version of the paper which appeared in *Hobbes Studies* 23.2 (2010), pp. 105–123!**

I would like to thank Richard Serjeantson as well as two anonymous referees for *Hobbes Studies* for their insightful comments on earlier versions of this essay.

## I. Introduction

Two of the perennial questions in Hobbes scholarship are whether the Laws of Nature, as laid out in chapters 16 and 17 of the *Elements of Law*,<sup>1</sup> in chapters 2 to 4 of *De Cive*,<sup>2</sup> in chapters 14 and 15 of *Leviathan*,<sup>3</sup> and in various other passages in these and other writings, *oblige* us, and if so, *in virtue of what* they do this.

At first, the puzzlement among Hobbes scholars may seem odd. After all, Thomas Hobbes makes it quite clear that by “Laws of Nature”, he means no more than “Conclusions, or Theorems concerning what conduceth to [our] conservation and defence”.<sup>4</sup> In light of this definition, it appears safe to say that for Hobbes, the only sense in which the Laws of Nature are obligatory is the familiar and unassuming sense in which all prudent courses of action are obligatory: if we decide against them, we are being stupid. What more, we may wonder, is there to ask?

Well, the scholarly questions lose their initial oddness as soon as we start to think about the function to which the Laws of Nature are put within *Leviathan’s* architecture. In *Leviathan*, Hobbes’s largest and most influential work of political philosophy, the Laws of Nature are supposed to provide the foundation for a very different kind of obligation, namely the kind of obligation incurred by “abandon[ing], or grant[ing] away [one’s] Right”.<sup>5</sup> It is in light of the commonly held belief that this latter kind of obligation – call it *contractual obligation* – can constitute a rather heavy burden, that Hobbes scholars have never ceased to wonder whether the obligatoriness of the Laws of Nature is really

---

1 Thomas Hobbes, *Elements of Law, Natural and Politic*, ed. J.C.A. Gaskin (Oxford: Oxford University Press, 2008)

2 Thomas Hobbes, *On the Citizen*, trans. M. Silverthorne, ed. R. Tuck (Cambridge: Cambridge University Press, 1998)

3 Thomas Hobbes, *Leviathan*, ed. R. Tuck (Cambridge: Cambridge University Press, 1996)

4 Hobbes, *Leviathan*, I.XV, 80 (page numbers in *Leviathan* follow the pagination of the 1651 edition); see also Hobbes, *On the Citizen*, II, 33.

5 Hobbes, *Leviathan*, I.XIV, 65; see also I.LXXI, 111.

just based on prudence, i.e. on properly reflected self-interest. And surely, this is no small question to ask. If the answer is yes, then either the common belief that honouring an obligation can go against reflected self-interest is false, or the *Leviathan* is flawed at its theoretical heart.

This paper aims at providing a new perspective on the issue by proposing that Hobbes is as unsatisfied with *Leviathan's* treatment of contractual obligation as many of his modern readers are. In fact, Hobbes's engagement with contractual obligation is a continuing struggle whose beginning long pre-dates the *Leviathan* and whose resolution – if that is what making peace with premises formerly considered below one's philosophical ambition amounts to – does not come until the very end of Hobbes's intellectual career. On this account, the real constant at the centre of Hobbes's changing philosophical system is the aim to furnish a secular “theodicy”, i.e. a reconciliation between man and the “Mortall God”<sup>6</sup> that is his commonwealth.<sup>7</sup> This paper will argue that if we acknowledge Hobbes's struggle in furnishing his secular theodicy instead of “reconstructing” his theory and finding in it more consistency than there is, we will be rewarded with a more historically accurate picture of the early modern intellectual scene, while giving up no part of our philosophical grasp of Hobbes's project.

The essay proceeds in four steps. The next – the second – section introduces Hobbes's early non-normative theory of obligation. The third section discusses his turn to a prudence-based conception of obligation. The fourth section shows that Hobbes remains unsatisfied with his own account, and the fifth section finally brings out some interesting consequences of Hobbes's theoretical unhappiness in his later writings.

---

6 Hobbes, *Leviathan*, I.XV, 80

7 With the exception of some sections paraphrasing or quoting Hobbes's own (decidedly gendered) opinions, I use “he” and “his” to refer to persons of all genders.

## II. Obligation as a mechanical phenomenon

Naturally, current Hobbes scholarship focusses on Hobbes's best-known work, the *Leviathan*. It is important to note, however, that before it, Hobbes had already completed several other books of political philosophy, most notably the *Elements of Law* and *De Cive*. Of course, these works parallel the *Leviathan* in important respects. Firstly, their principal goal is to counsel their readers to refrain from revolting against the powers that be – who or whatever happens to play this role.<sup>8</sup> Secondly, their counsel rests on the thesis that any political arrangement with a powerful sovereign is better than the state of nature. Finally, they argue that the obligation characteristic of such arrangements – political obligation – can be accounted for in terms of contractual obligation.<sup>9</sup> However, there is a difference which turns out to reflect an important shift in the association between contracts and the Laws of Nature in Hobbes's developing philosophy.

What sets the *Elements* and *De Cive* apart from Hobbes's later philosophy is that these books are much more occupied with the *making* of contracts than with their *keeping*. And there is an interesting reason for it: the early Hobbes thinks that once an agent *wills* a contract, he thereby *mechanically, i.e. physically, necessitates himself* to do what the contract requires of him, or as Stephen Darwall (who has pointed out this peculiarity of Hobbes's early thought) puts it: “contract obligates by restricting the liberty of the contracting agent.”<sup>10</sup> It is in the context of his mechanistic account of the will that

8 I follow Quentin Skinner in interpreting Hobbes as arguing for the submission to who or whatever is the *de facto* holder of sovereign power; see Quentin Skinner, “Conquest and Consent”, in his: *Visions of Politics* (Cambridge: Cambridge University Press, 2002). It seems to me that this interpretation is now supported by the majority of Hobbes scholars, with interpretive controversy focusing mainly on details surrounding transitional periods; for a fine discussion of the interpretive landscape, see Kinch Hoekstra, “The *de facto* Turn in Hobbes's Political Philosophy”, in T. Sorrell and L. Foisneau, *Leviathan after 350 years* (Oxford: Oxford University Press, 2004), 33-74.

9 Although, of course, this does not mean that the obligation owed to the sovereign *is* contractual obligation. For Hobbes, individuals strike a social contract to institute a sovereign who is himself not a party to the contract. See *Leviathan*, I.XVII, esp. 87f.

10 Stephen Darwall, *The British Moralists and the Internal 'Ought' 1640-1740* (Cambridge: Cambridge

Hobbes frames the connection between contracts and the Laws of Nature. In order to appreciate the early Hobbesian position and thus the original role of the Laws of Nature, we must go back to his early theory of deliberation, will and freedom, or – as we can also put it – to the beginning of Hobbes’s struggle with contractual obligation.

Already early on, Hobbes had been a proponent of the doctrine of determinism. In his mechanistic view of the world, consisting exclusively of matter in motion, human agents are, like everything else, fully governed by the natural, i.e. physical, laws. Human deliberation, for Hobbes, is a “weighing up the advantages and disadvantages of the action we are addressing (as on a pair of scales), where the weightier consideration necessarily goes into effect by its own natural inclination”.<sup>11</sup> Consequently, the will is not conceived as a faculty of the mind, as mainstream scholasticism would have it, but merely as the outcome of deliberation, or in Hobbes’s own words, as “the last Appetite in Deliberating”.<sup>12</sup> In fact, if there is anything like a “free will”, this is only because “freedom”, or “liberty”, for Hobbes, consists “simply [in] the absence of obstacles to motion”.<sup>13</sup> If we now add the innocent-sounding claim that “by agreeing to a future action [one] wills that it be done,”<sup>14</sup> we get a doctrine of decision and action which can account for the keeping of contracts rather straightforwardly. Since “obligation begins where liberty ends”,<sup>15</sup> the keeping of contracts is a matter of the physically necessary course of the world. Hobbes sometimes calls this state of being necessitated to do something a “natural obligation”.<sup>16</sup>

---

University Press, 1995), ch. 3, 65

11 Thomas Hobbes, *On the Citizen*, XIII, 152

12 Hobbes, *Leviathan*, I.VI, 28

13 Hobbes, *On the Citizen*, IX, 111

14 Hobbes, *On the Citizen*, III, 45

15 Hobbes, *On the Citizen*, II, 36

16 Hobbes, *On the Citizen*, XV, 174

Of course, a fully mechanistic theory of the will has rather significant problems. Not only is it unclear whether it leaves room for *arguments* for or against particular courses of action – arguments such as those put forward by Hobbes himself. It is also unclear whether someone who assumes a fully mechanistic theory of the will can so much as make *sense* of the notion of commitment (to a contract, to a norm, even to a goal), as commitment seems to presuppose liberty of a more substantial kind than mechanism makes room for. Although Hobbes did not – and could not – have a fully worked out philosophical treatment of these problems, he did have a feel for the looming trouble. His sense of trouble can plausibly be seen at work in his early account of freedom. While for the later Hobbes, liberty “properly signified” is only the absence of *external* impediments to motion,<sup>17</sup> for the Hobbes of the *Elements* and *De Cive*, there are also such things as “*discretionary*” impediments to motion, i.e. chosen impediments (the Latin original has “*impedimenta arbitraria*”<sup>18</sup>). According to this view, an agent is free as long as his will is not (yet) fixed, i.e. as long as he is (still) in the process of deliberating. Conversely, he is unfree once he has fixed his will. The agent – for that is what *chosen* impediments imply: agency – is then (only then) irresistibly necessitated. It is not implausible to see this philosophical twist as a reaction to the need for an account of the role of arguments in deliberation as well as the phenomenon of commitment – a reaction, moreover, which leaves intact the idea that to will is to physically necessitate one’s own conduct.

Now, whether or not Hobbes’s early mechanistic account of deliberation is defensible, it does – as I mentioned above – have the consequence of focussing his entire philosophical project on the *making* of contracts rather than on their *keeping*. Since the

---

17 See Hobbes, *Leviathan*, I.XXI, 107.

18 *Thomas Hobbes Opera Philosophica Quae Latina Scripsit Omnia*, ed. W. Molesworth, (London: Routledge/Thoemmes, 1839), vol. II, 259; Hobbes, *On the Citizen*, IX, 111

latter is just an aspect of the mechanical course of the world, the only problem is to get men genuinely to *will* the contracts which are necessary for peace and thus self-preservation. Not only are many people ignorant about how best to achieve self-preservation (hence the need for Hobbes's books), there is also the evident possibility of free-riders who only *pretend* to will the appropriate contracts. In fact, in the *Elements* and in *De Cive*, it is *this* problem which necessitates the installation of a Hobbesian sovereign. Here, the sovereign's job is most plausibly interpreted as ensuring that individuals' vows are expressions of their wills, rather than ensuring that willed contracts are kept. This interpretation, by the way, fits nicely with François Tricaud's observation that in *De Cive*, the state of nature is characterised by the absence of contracts and only subsidiarily by the absence of a sovereign, whereas the emphasis is reversed in *Leviathan*.<sup>19</sup>

Let us finally apply this discussion to the Laws of Nature. What is important to understand is that Hobbes (late *and* early) introduces the Laws of Nature with two aims in view. Firstly, they are to codify what is requisite for self-preservation and hence in our interest. According to *De Cive*, the Laws of Nature are "Dictate[s] of right Reason, conversant about those things which are either to be done, or omitted for the constant preservation of Life, and Members, as much as in us lyes."<sup>20</sup> Secondly, their subject matter is to centre on contracts, as the titles and the contents of the chapters outlining the Laws of Nature clearly show.<sup>21</sup> If we now take into account the mechanistic account

---

19 See François Tricaud, "Hobbes's Conception of the State of Nature from 1640 to 1651: Evolution and Ambiguities", in G.A.J. Rogers and A. Ryan (eds.), *Perspectives on Thomas Hobbes* (Oxford: Clarendon Press, 1988), 110.

20 Hobbes, *On the Citizen*, II, 33. *Leviathan's* corresponding passage defines the Laws of Nature as "Conclusions, or Theorems concerning what conduceth to (our) conservation and defence"; see Hobbes, *Leviathan*, XIV, 65.

21 *De Cive's* first chapter on the Laws of Nature is entitled "On the natural law of contracts"; *Leviathan's* central chapter XIV is entitled "Of the first and second Naturall Laws, and of Contract". Note also that where these chapters on the Laws of Nature are not about contracts, they are about *contract-like* self-restraint such as "Gratitude", "Mutall accommodation" and the mutual acceptance of certain rules of arbitration. See *De Cive*, ch. II and III; *Leviathan*, ch. XIV and XV.

of the will, not only must the second aim be interpreted as dealing only with the *making* of contracts, we also get the consequence that although the Laws of Nature are normative in that we have *reason* to follow them, but are *able* to violate them (at least, that seems to be the point of Hobbes's thesis that an agent's conduct is physically necessitated only *once* his will is fixed), there is *no temptation* to break them. Since the Laws of Nature only prescribe contracts that are prudent *anyway*, and since the keeping of these contracts is a purely mechanical affair, the Laws cannot ever be a *burden*.

### **III. Hobbes's turn to self-interest**

Although the theory has the desired output, it is quite clear that around the time Hobbes sets to work on his *Leviathan*, he is no longer satisfied with his early treatment of contracts. In fact, in the *Leviathan*, there is no more mention of "natural obligation". Instead, *contractual obligation* becomes an openly normative term, which is to say that Hobbes acknowledges the physical possibility of the breach of a genuinely willed contract, and instead speaks of the *wrongness* of such breach. As we will see, this view represents the next stage in Hobbes's long struggle with contractual obligation.

Before tackling *Leviathan's* new official doctrine, let us think about what made Hobbes replace his project's philosophical engine. Admittedly, the evidence is somewhat sparse, but we can identify three reasons for the replacement. Firstly, Hobbes comes to see some manifest problems with his earlier view of freedom and necessitation. It is not just from today's vantage point that it is hard to make sense of the view that an agent in the process of deliberation is any more free than an agent who has concluded his deliberation. In Hobbes's correspondence with Bishop Bramhall on liberty and necessity,<sup>22</sup> Bramhall points out to Hobbes that this view leads to obvious absurdities:

---

<sup>22</sup> Thomas Hobbes and John Bramhall, *Hobbes and Bramhall on Liberty and Necessity*, ed. V. Chapell (Cambridge: Cambridge University Press, 1999), xxii



the man, for example, who “deliberates whether he shall play at tennis, and at the same time, the door of the tennis court is fast locked against him”, would count as both free and unfree to play tennis.<sup>23</sup> In Hobbes’s mature theory of liberty – although arguably still lacking in theoretical rigour<sup>24</sup> – liberty “properly signified” is thus strictly the absence of *external* impediments, and there is much more insistence on the compatibility between freedom and necessity.<sup>25</sup>

Secondly, Hobbes realises that his early treatment of deliberation and contracts leaves no room for the *voiding* of contracts, and that this is a problem. In fact, Hobbes’s commitment to the physically necessary connection between the will to strike a contract and the conduct demanded by it already gets tested in *De Cive*, when, in a revealing passage just pages away from the introduction of the Laws of Nature, Hobbes wants to affirm the common opinion that contracts violating the Laws of Nature (“anything illicit”) do not bind *even if sincerely willed*.<sup>26</sup> One is also reminded of the passages in which Hobbes writes that where free-riding is prevalent, the honouring of one’s own vows is obligatory only “in foro interno”.<sup>27</sup> These ideas clearly demand a basis different from the thesis that to will a contract is to physically necessitate oneself.

We may want to add a more speculative, but plausible, third consideration. It is not unlikely that Hobbes comes to see that the fact that his project is not just *hypothetical* forces him to replace his theory of contractual obligation. It is quite clear that Hobbes is committed to the thesis that *we* – his readers who live in civil society – are bound by a social contract. He thinks that *our* sovereign was in fact instituted by way of such a

---

23 Hobbes, *English Works*, vol. II, 346

24 See Quentin Skinner, “Thomas Hobbes on the Proper Signification of *Liberty*”, in: *Transactions of the Royal Historical Society* 40 (1990), 121-51.

25 See, for example, Hobbes, *Leviathan*, I.XXI, 108ff.

26 See Hobbes, *On the Citizen*, II, 37.

27 Hobbes, *On the Citizen*, III, 54; Hobbes, *Leviathan*, I.XV, 79. Note that the talk of an obligation in the internal court is problematic not only on *De Cive*’s doctrine of contracts, but also on *Leviathan*’s.

contract. Since this implies the presence of a *will* to contract (even if it is a will whose original expression is long past), the only way to maintain that there is *still* a danger of revolt and war – and Hobbes very obviously believes in this danger – is to give up the idea that the willing of a contract leads inevitably to its keeping.

In short, Hobbes sees more than one reason to replace his old doctrine of contracts with a new one. With the absence of “natural obligation” in *Leviathan*, contractual obligation becomes a normative notion, and since it is natural for Hobbes to think that for a phenomenon to be normative is for it to be covered by the Laws of Nature, the latter now take on the job of grounding the *keeping* of contracts (while also keeping their old job of grounding the *making* of certain contracts). At this point, however, it is crucial to remember that the Laws of Nature are “Conclusions, or Theorems concerning what conduceth to [our] conservation and defence.”<sup>28</sup> What this means is that Hobbes’s analysis of normativity (which never played that prominent a role in his early work) now forces him to say that the wrongness of contract-breaking is the wrongness of imprudent courses of action.

This raises an important worry – namely the worry that Hobbes’s conception of normativity is too weak to carry the weight he now places on it. The danger lies in the common belief that it *is* possible that a contract obliges its parties to act *against* their current, reflected, self-interest, even if the making of the contract was prudent.<sup>29</sup> In fact,

---

28 Hobbes, *Leviathan*, I.XV, 80. Note, by the way, that the wording (“theorems”, “conclusions”) suggests that the Hobbes of the *Leviathan* is quite attentive to the fact that instrumental reasoning is non-trivial – and thus, presumably, fallible. This interpretation fits well with Hobbes’s repeated reminders of the harm that bad thinking and bad philosophy can do in the political realm; see Hobbes, *Leviathan*, ch. 5, 18ff. This chapter is in stark contrast to the *Elements*’ fifth chapter with its dismissive treatment of logic and its unwillingness to discuss the dry matter of ratiocination. *De Cive* omits a treatment of the quality of ratiocination or philosophy altogether.

29 Note that the common opinion is not just that honouring a (prudently struck) contract can fall short of satisfying our *present desires*. It is that honouring a (prudently struck) contract can fall short of optimally achieving our long-term interests, even if the latter are understood as (counterfactually) cleared of inconsistencies and misinformation – hence the word “reflected” in my formulation. In fact, if it weren’t for the paradoxical sound, the thesis would be best put as saying that honouring a

since our notion of obligation is closely allied with our notion of *reason*, the danger lies in the thought that *pace Hobbes*, we can have *reasons* to act against our current, reflected, self-interest.

In the light of this problem, our philosophical options seem to be the following. Firstly, we can accept the thought that honouring an obligation stemming from a prudently made contract can go against reflected self-interest, and thus reject the thesis that prudence is the sole basis of contractual obligation. This can be put in different ways, depending on whether we are still prepared to adhere to Hobbes's definition of the Laws of Nature as theorems of prudence. Going with Hobbes's terminology, we would say that the Laws of Nature are not the basis of contractual obligation. If, on the other hand, we do not mind violating Hobbes's terminology but aim to keep the association between the Laws of Nature and contractual obligation in place, we would say that the Laws of Nature themselves must have some deeper normative foundation.

The second option is to stick to the thesis that prudence is all that is needed to account for contractual obligation. If we want to take this route, we must find a way to argue against the claim that honouring an obligation arising from a prudently made contract can violate reflected self-interest – and, ideally, offer some explanation of the claim's initial appeal. On Hobbes's terminology, this would amount to showing that honouring prudently struck contracts indeed never violates the Laws of Nature. (I take it that those who decide to take this second philosophical path are satisfied with Hobbes's terminology.)<sup>30</sup>

---

(prudently struck) contract can be *imprudent* (although it can be right).

30 It seems to me that the recent controversy over whether (or in what sense) the Laws of Nature are conditional on contingent aspects of human nature shows fundamental philosophical differences along the lines sketched here. See John Deigh, "Reason and Ethics in Hobbes's *Leviathan*", in *Journal of the History of Philosophy* 34 (1996); James Murphy, "Desire and Ethics in Hobbes's *Leviathan*: A response to Professor Deigh", in *Journal of the History of Philosophy* 38.2 (2000); Kinch Hoekstra, "Hobbes on Law, Nature and Reason" in *Journal of the History of Philosophy* 41.2 (2003).

Hobbes is acutely aware of the problem. In fact, *Leviathan's* central passages are most plausibly interpreted as a reaction to it. More specifically, in the 15<sup>th</sup> chapter, Hobbes sets himself the task of taking and defending the second of the two routes just outlined: Hobbes wants to show that once a contract has been struck (which, in the case of a rational and adequately informed agent, implies that the striking of the contract was prudent), there *must* be conclusive reasons of self-interest to keep it – i.e. its keeping must be implied by the Laws of Nature. If the argument is successful, it establishes that the Laws of Nature are no more a “burden” in *Leviathan* than they are in *De Cive*. Or, in Hobbes’s own words: it establishes that they are “eternall; *and yet easie*”.<sup>31</sup>

To see that this is indeed Hobbes’s goal, and to see how Hobbes attempts to reach it, we must turn to the famous reply to the Foole in the first half of chapter 15 of *Leviathan*. This has been a major focus of attention in recent literature on the status of the Laws of Nature, and in this, we will follow that body of literature. We will diverge from (most of) it, however, in that we will read the passage above all with a view to understanding the *difficulties* facing – and ultimately overpowering – Hobbes. Admittedly, this is not the easiest of tasks, for Thomas Hobbes is not the kind of author who confesses that he is unsatisfied with a formulation or struggling with a philosophical problem. However, there are highly interesting clues to work with.

In his reply to the Foole, Hobbes aims for the conclusion that it is always rational, i.e. prudent, to keep covenants. Covenants are contracts wherein one party performs immediately, the other later. Covenants are of course the kinds of contracts which, for one thing, are most liable to breach, and which, for another, do the most work in

---

31 Hobbes, *Leviathan*, I.XV, 79, side note, emphasis added. I actually think that this side note constitutes an important piece of evidence not yet (as far as I am aware) picked up by Hobbes’s commentators. Note that “easy”, on my reading, signifies “never in breach of reflected self-interest”, *not* “never in breach of present desires”.

Hobbes's theory of the institution of the sovereign. Since Hobbes defines justice as the keeping of covenants<sup>32</sup> immediately before the Foole enters, we can agree that in his reply to the Foole, Hobbes argues for the rationality of justice. But first, enter the Foole.

The Foole hath sayd in his heart, that there is no such thing as Justice, and sometimes also with his tongue; seriously alleaging, that every mans conservation, and contentment, being committed to his own care, there could be no reason, that every man might not do what he thought conduced thereunto: and therefore also to make, or not make; keep, or not keep Covenants, was not against Reason when it conduced to ones benefit. ... [Y]ou may call it injustice, or by what other name you will; yet it can never be against Reason, seeing all the voluntary actions of men tend to the benefit of themselves; and those actions are most Reasonable, that conduce most to their ends.<sup>33</sup>

In spite of the striking similarity in character and style between Hobbes and the Foole<sup>34</sup> and in spite of the apparently shared premise about the nature of reason, the former writes that the latter's "specious reasoning is nevertheless false."<sup>35</sup> Why? What is Hobbes's argument?

Hobbes begins his reply by going through various cases to which the Foole's reasoning might be thought to apply. Firstly, he considers the case of "mutual promises" in the state of nature, i.e. a mutual exchange of promises of future action. The second case is a covenant in the state of nature in which the other party has already performed. The third

---

32 See Hobbes, *Leviathan*, I.XV, 71.

33 Hobbes, *Leviathan*, I.XV, 72

34 See Ross Harrison, *Hobbes, Locke, and Confusion's Masterpiece* (Cambridge: Cambridge University Press, 2003), ch. 4, 121.

35 Hobbes, *Leviathan*, I.XV, 72

is a covenant where the other party is to perform second, but “there is a Power to make him performe”<sup>36</sup>. Now, the first case is immediately brushed aside, as it does not fall under the definition of covenant. This leaves the second and the third case up for consideration. Although there is no explicit argument, almost the entire discussion that ensues concerns the second case. Hobbes finds the third case self-evident, and this should not come as a surprise, as it is clear that where there is a “Power to make [the other party] performe”, there is no big question about the rationality of performing now.

It is the second case, then, which for Hobbes is the important case to discuss. Before we look at his discussion, let us pause for a short moment and note that not only does this make the order of the three cases quite puzzling from a strategic point of view; going through the three cases *at all* is a strange manoeuvre, as it is *so evidently* the second case for which the Foole has proposed breach as a rational course of action. It is only the second case (and those instances of the first case which turn out to be describable as instances of the second case, i.e. those where there is a time-difference between the agreed performances) which poses the problem of rational performance at all.

At any rate, Hobbes’s argument to the effect that non-compliance in the second case is irrational appears to have two parts, a preliminary part and a main part. Firstly, he writes, if a course of action cannot be expected to be advantageous, but turns out to be so because of some “unforeseeable”, improbable, event, then the action cannot count as “rational”.<sup>37</sup> An action’s rationality depends on the legitimate *expectation* of its outcome, not the outcome itself. This point, of course, does not amount to an argument for the irrationality of non-performance. For that, the further premise that non-performance cannot indeed be expected to be advantageous is needed. In fact, since

---

36 Hobbes, *Leviathan*, I.XV, 73

37 Hobbes, *Leviathan*, I.XV, 73

Hobbes needs to show that non-performance is *never* rational, the argument would have to rely on the premise that it can *never* be expected to be advantageous. Let us see whether Hobbes goes on to supply it in the second part of his argumentative strategy.

Secondly, that in a condition of war wherein every man to every man (for want of a common power to keep them all in awe) is an enemy, there is no man can hope by his own strength or wit to defend himself from destruction without the help of Confederates ... and therefore, he which declares he thinks it reason to deceive those that help him, can in reason expect no other means of safety, than what can be had from his own single power. He, therefore, that breaketh his covenant, and consequently declareth that he thinks he may with reason do so, cannot be received into any Society, that unite themselves for Peace and Defence, but by the error of them that receive him ... and therefore he perisheth ... and consequently [he has acted] against the reason of his preservation ...<sup>38</sup>

It is quite safe to say that this answer has not convinced many serious readers. In fact, one is tempted to exclaim that “the problems with it are so astoundingly obvious that one must wonder how Hobbes dared to give it.”<sup>39</sup> Since the answer applies only to cases in which the (prospective) contract-breaker is publicly identifiable, it begs the most important question and thus effectively leaves the Foole unrefuted. The Foole *obviously* accepts that it is *normally* rational to perform. In the light of Hobbes’s answer, the Foole’s point could easily be re-formulated as stating that *sometimes*, in the special case in which the contract-breaker *can* expect to get away without being detected – all the cautionary points of Hobbes’s answer considered – his non-performance *can* be

---

38 Hobbes, *Leviathan*, I.XV, 73

39 A. Zaitchik, “Hobbes’s Reply to the Foole: The Problem of Consent and Obligation”, in *Political Theory* 10.2 (1982), 246

expected to “conduce to [his] benefit” and hence be rational. On this, Hobbes has *nothing* to say.

This remarkable mismatch between the problem and Hobbes’s reaction, which is by no means cleared in the subsequent passages, has received three kinds of response. The first acknowledges Hobbes’s problems in grounding obligation on prudence and take them as an indicator of the tacit role of some other source of obligation in Hobbes’s thought. Going back to my sketch of the two ways of dealing with the claim that honouring an obligation can go against self-interest, this would be to take the first line. The second response attempts to rescue Hobbes’s answer by reading one of various supporting premises into Hobbes argument, while maintaining the thesis that in the end, obligation boils down to properly reflected self-interest. The third response claims that Hobbes’s reply to the Foole is not actually meant to deal with the theoretical problem arising from the belief that obligation stemming from a prudently made contract can be a burden, and that Hobbes’s argumentative aim in his reply to the Foole is much more modest. Let me briefly give a few examples of the three classical responses, and then consider the merits of acknowledging Hobbes’s failure instead of coming to his rescue.

An example of the first response is Aloysius Martinich’s. Martinich thinks that Hobbes’s argument for the correctness of honouring obligations contains a premise about divine command, and holds that the obligatoriness characteristic of the latter is of a different sort from that of prudence.<sup>40</sup> According to Martinich, the Laws of Nature are morally obligatory because they are God’s word; the fact that they are *also* rational in the sense

---

<sup>40</sup> See Aloysius Martinich, *Two Gods of Leviathan: Thomas Hobbes on Religion and Politics* (Cambridge: Cambridge University Press, 1992), see also Geoff LeBuffe, “Hobbes on the Origin of Obligation”, in: *British Journal for the History of Philosophy* 11.1 (2003), 17.



of guiding us towards self-preservation is logically unconnected with their moral obligatoriness.<sup>41</sup>

The second response was and is to accept Hobbes's aim of building contractual obligation on prudence alone, and to look for supporting premises to help him make his case against the Foole. In fact, there are quite a lot of responses along these lines. Stephen Darwall argues that contracting parties in the state of nature are not in a situation of risk, but in one of uncertainty, and that this consideration – along with common knowledge about it – supports compliance.<sup>42</sup> Gregory Kavka thinks that Hobbesian agents do not primarily decide about how to *act*, but about what “*rules*” of action to adopt, and further that there is a principle of “disaster-avoidance” at work, according to which agents must discount interest-calculations in situations in which disasters are possible.<sup>43</sup> Ross Harrison locates the covenant situation somewhere *between* the state of nature and civil society, in a stage of “state formation”, in which special rules of reasonable conduct obtain,<sup>44</sup> and David Gauthier argues that Hobbes had access to the argument that committing oneself to a choice that is suboptimal when considered only in view of the range of immediate alternatives *can* be optimal given a second (or more) strategic agent(s) with known beliefs and preferences.<sup>45</sup>

---

41 See Martinich, *Two Gods*, especially pp. 136ff. Note, by the way, that in spite of first appearances, Howard Warrender's take on the *Leviathan* does not fall into the category at hand. Although Warrender is correctly associated with the divine-command-hypothesis, he does seem to be happy with the doctrine that the Laws of Nature rest on prudence alone: for him, the point is just that a prudent course of action aims not merely at survival in this world, but also and crucially at well-being in the promised afterlife. See Howard Warrender, *The Political Philosophy of Hobbes: His Theory of Obligation* (Oxford: Oxford University Press, 1957), in particular part III, where he writes that for Hobbes, someone who does not believe in God and an afterlife is not obligated to obey the Laws of Nature, and may even have a reason (presumably a subjective reason) to defy them.

42 See Darwall, *The British Moralists and the Internal 'Ought'*, 77

43 See Gregory Kavka, *Hobbesian Moral and Political Theory* (Princeton: Princeton University Press, 1986), 338ff.

44 See Ross Harrison, *Hobbes, Locke, and Confusion's Masterpiece* (Cambridge: Cambridge University Press, 2003), ch. 4, 123

45 See David Gauthier, “Hobbes's Social Contract”, in G. A. J. Rogers and A. Ryan (eds.), *Perspectives on Thomas Hobbes* (Oxford: Clarendon Press, 1988), 143ff.

The third response is to say that the impression of Hobbes's spectacular failure is simply due to our exaggerated expectations. Kinch Hoekstra has recently argued for this thesis,<sup>46</sup> claiming that what Hobbes really wants to show is just the irrationality of openly *declaring* one's intent to break covenants. In fact, Hoekstra thinks that Hobbes's thesis also covers acts which are so conspicuously in contempt of justice (i.e. of covenant-keeping) as to be quasi-declaratory. On this reading, the Foole mainly argues for the exploitation of the (supposedly) wide-spread error of admitting covenant-breakers into society, and Hobbes's reply is that while this error may indeed occur, it is never a good idea to *rely* on it. In other words, Hoekstra takes Hobbes's reminder that an action's rationality depends not on its de facto consequences, but on the legitimate *expectation* of its consequences, as a central point in the argumentative strategy.

The three kinds of response clearly have philosophical merits, but none of them adequately reflects *Hobbes's own* thinking, including its twists and turns. In order to show this, I want to offer a very different reaction to Hobbes's reply to the Foole by gauging Hobbes's own estimate of his answer's degree of success. This kind of analysis, I think, can show where all three extant responses go wrong, namely in their implication that Hobbes is satisfied with his reply to the Foole.

---

46 See Kinch Hoekstra, "Hobbes and the Foole", in: *Political Theory* 25.5 (1997), 620-654

#### IV. Hobbes's doubts about his official doctrine

Interestingly, a careful reading not just of the English *Leviathan* of 1651, but also of the Latin *Leviathan*, published some 17 years later, reveals that Hobbes *himself* is unhappy with his reply to the Foole. This does not mean that he tacitly relies on a source of obligation different from prudence, or that he does not really care to show whether or how contractual obligation flows from prudence. All it means is that there are philosophical problems in Hobbes's arguments, and that he is painfully aware of them.

Firstly, then, in both editions of *Leviathan*, Hobbes falls back on uses of the terms "justice" and "injustice" which sit very oddly with his reply to the Foole, no matter which of the available interpretations we work with. This falling back parallels another of Hobbes's slips which we find in his problematic talk about the "internal" and the "external court", whose oddness also does not depend on a specific, contentious, view of Hobbes's aims in his reply to the Foole. Secondly, in the Latin *Leviathan*, the key passage of the reply to the Foole is significantly shorter than the passage in the earlier English text – and tellingly different.

Let us start with the first observation. There are passages which employ a concept of justice quite at odds with the one which informs the answer to the Foole. Justice, remember, is integral to our study as it is defined by Hobbes as the keeping of covenants. An example of the kind of slip of which I accuse Hobbes, now, is a passage immediately after the attempt at rebutting the Foole, in which Hobbes explains that just because of a single "just" act, a generally "Unrighteous man" will not be considered just.<sup>47</sup> Although this point does not seem to stand in need of an explanation, Hobbes gives one – and a remarkably problematic one at that. He writes that the man will

---

<sup>47</sup> Hobbes, *Leviathan*, I.XV, 74

continue to count as unjust “because his Will is not framed by Justice, but by the apparent benefit of what he is to do”.<sup>48</sup> Here, Hobbes effectively says that in order for an act to be just, it needs to be done for the *right reason* – regard for justice – and that this right reason is something *other* than the act’s expected benefit. This is odd indeed, as it is not only hard to square with Hobbes’s very doctrine that contractual obligation is founded on self-interest, but also goes against Hobbes’s official definition of injustice as merely “the *not Performance of Covenant*.”<sup>49</sup>

It might be responded that the context of the quote is a discussion of the difference between the concept of justice as applied to persons and the concept of justice as applied to singular actions, and that the discussion just aims at establishing that robust evidence about the former can outweigh the evidence of a single instance of the latter. But this sort of thing surely does not call for the words Hobbes finds to describe a “Will ... framed by Justice”: “[t]hat which gives to human actions the relish of justice, is a certain nobleness or gallantness of courage, rarely found, by which a man scorns to be beholden for the contentment of his life, to fraud, or breach of promise.”<sup>50</sup> If *this* just means that through just actions, the just man looks for his own benefit, which (as it were) often comes via reputation, then we have no departure from Hobbes’s official position. But then why make a difference between looking for “one’s apparent benefit” and having a will “framed by Justice”?

Now, this – it might be responded – just shows that Hobbes means to distinguish between acts and *rules* of action. And surely, something like this is part of Hobbes’s point here. But if the reason for which an unjust man continues to count as unjust is because in his single just act, he aims for his own benefit (even “apparent” benefit), then

---

48 Hobbes, *Leviathan*, I.XV, 74

49 Hobbes, *Leviathan*, I.XV, 71, emphasis in original

50 Hobbes, *Leviathan*, I.XV, 74

the ordinary just man surely also deserves to count as unjust, which is not what Hobbes wants to say. It is clear that in the passages under scrutiny, Hobbes is rather under the influence of a backward-looking notion of justice – a sort of justice for which there is a reason *over and above* the forward-looking considerations of prudence.<sup>51</sup> Surely, Hobbes is hardly entitled to a position of this sort.

So *is* there, for Hobbes, a genuine backward-looking reason for the keeping of covenants besides rational self-interest? Well, that is unclear. Only a few pages further on, when he discusses *equity*, which is nothing but distributive justice,<sup>52</sup> Hobbes goes back to his official defence of justice (“He therefore that is partial in judgement ... is the cause of war”<sup>53</sup>). It is not a change in the notion of “justice”, nor in its defence, it is rather the *temptation* to depart from the official line to which I want to draw attention. Another example of this temptation can be discerned in Hobbes’s thesis, propounded in the context of his problematic distinction between the “internal” and the “external court”,<sup>54</sup> that even an action *in accord* with the Laws of Nature can constitute a breach of the Laws of Nature, if the agent’s “*purpose* was against the law.”<sup>55</sup> Like the “will framed by justice”, this talk is alien to the rest of Hobbes’s philosophical system, and it so manifestly leads to problems that it can only be interpreted as a slip of the tongue revealing Hobbes’s own lack of faith in his master argument.

Secondly, as I said, the Latin *Leviathan* of 1668 incorporates some suspicious changes. It is not only that in almost no other chapter in Books I and II of *Leviathan*, Hobbes so carefully revises his formulations as he does in chapter 15; indeed, one of the altered

---

51 I borrow this terminology from Martin Hollis, “Penny Pinching and Backward Induction”, in: *The Journal of Philosophy* 28 (1991), 473-488, esp. 488.

52 See Hobbes, *Leviathan*, I.XV, 77.

53 Hobbes, *Leviathan*, I.XV, 77

54 See Hobbes, *Leviathan*, I.XV, 79.

55 Hobbes, *Leviathan*, I.XV, 79, emphasis added

passages is the very reply to the Foole. The main change is the removal of the lines suggesting that the Covenant-breaker operates *openly* “and consequently *declareth* that he may with reason do so”.<sup>56</sup> This might seem like the removal of an innocent source of misunderstanding, but as a matter of fact, it mirrors Hobbes’s problem with the fact that the purported irrationality, i.e. imprudence, of Covenant-breaking rests *exclusively* on the latter’s liability to being found out. We must remember that the only serious part of Hobbes’s attempt at refuting the Foole’s thesis was the possibility that the covenant-breaker could be identified as such and excluded from further social co-operation, including further schemes of non-aggression. The suggestion, in the English *Leviathan*, that the covenant-breaker *presents himself* as someone who considers it rational to cheat on those co-operating with him puts the finger right on the weakest link of Hobbes’s argument. So, Hobbes concludes, it has to go.

It might be asked what reason there is for preferring an attribution of such an intellectual dishonesty to Hobbes over other possible interpretations. The reason lies in the remarkable fact that Hobbes never even *considers*, let alone answers, the obvious question “What if the covenant-breaker is *safe*?”. Given that Hobbes’s general philosophical aims, which have not changed since the *Elements*, clearly require an answer to this question (which has become salient with the abandonment of the idea of “natural obligation” after *De Cive*), and given that the *Latin Leviathan* does not actually supply anything like a *clarification* of the matter, the most plausible explanation of Hobbes’s silence is that he genuinely does not know how to cope with the possibility of situations where the covenant-breaker *is safe*. He realises that the intellectual resources

---

<sup>56</sup> Thomas Hobbes, *Leviathan*, with selected variants from the Latin edition of 1668, ed. E. Curley (Indianapolis: Hackett, 1994), 92, footnote 7, emphasis added. Kinch Hoekstra has also noted the absence of the relevant lines in the Latin *Leviathan* (“Hobbes and the Foole”, 626). Hoekstra, however, who is in the business of reconstructing Hobbes, merely takes this to be evidence for a slight modification of Hobbes’s generally modest thesis. Remember that for Hoekstra, Hobbes only argues against the advisability of coming across as someone who considers it rational to break contracts.

of the discourse of *prudence* will not do the job of supplying a plausible forward-looking reason for the keeping of contracts, and an *ad-hoc* empirical thesis about the inevitability of the identification of anti-social elements in human societies is far below Hobbes's philosophical ambition (i.e. far too strong). Hobbes has no good answer and knows it, and it is for this reason that from time to time in the *Leviathan*, we come across scattered remarks about the role of God, about justice for its own sake and about all kinds of auxiliary grounds for the argument to stand on. I do not want to go so far as to account for the whole of Book II of *Leviathan* in this way, but I do want to suggest that Hobbes's awareness of the dysfunctions of the theoretical core of his *Leviathan* explains much of his straying from the official line of his work.

## V. Concluding remarks

In fact, this reading of Hobbes's development actually sheds some light on certain puzzling facts about the direction of Hobbes's thought *after* his *Leviathan*.

Let me sketch the further development of Hobbes's thought. Hobbes's work, as I have indicated at the outset, is above all a "theodicy", a reconciliation between man and the "Mortall God" that is his commonwealth. After reading it, Hobbes hopes, "you [the reader] will think it better to enjoy your present state (though it may not be the best) rather than go to war, and after you have been killed or died of old age, leave other men in other times to have a better life".<sup>57</sup> Now, if Hobbes really did grow unhappy with his arguments for the making and keeping of certain covenants – which obviously played the key role in Hobbes's theodicy in *Leviathan* – then it is a compelling possibility that after *Leviathan*, he was looking for new arguments for his theodicy project. In *Behemoth*, it seems, Hobbes has not quite found them yet, for his only answer to the

---

<sup>57</sup> Hobbes, *On the Citizen*, Preface to the Reader, 14

question whether tyrants must be obeyed, even when they demand of someone “with [his] own hands to execute [his] father,” is that “[w]e never have read nor heard of any King or tyrant so inhuman as to command it”.<sup>58</sup> But this kind of reply is certainly at odds with *Leviathan*, which would just have answered “yes”. Already in *Behemoth*, we can see that Hobbes is less resistant to considering the possibility of tyranny worse than the state of nature, hoping that “God forbid that so horrible, unchristian, and inhuman a design should ever enter into the King’s heart”.<sup>59</sup> But it is not before the *Dialogue between a Philosopher and a Student, of the Common Laws of England* that Hobbes starts to exhibit new modes of thinking after his failure to rebut the Foole.

Hobbes now propagates the sovereign’s regular consultation of the parliament,<sup>60</sup> calls the waging of war without consultation of military experts a *sin*,<sup>61</sup> demands of the parliament to refuse to back policy if it is against the law of *salus populi*<sup>62</sup> and entertains very new ideas of reason as something like impartiality rather than as a “reckoning” in the service of self-interest – complaining even that “the greatest part of men are so unreasonable, and so *partial to themselves*”!<sup>63</sup> These points are particularly striking, as the *Dialogue’s* general philosophical drive is to argue against the encroachment of the sovereign’s legal powers by the growing body and practice of common law. The *Dialogue* makes it clear that if common law is an expression of reason – a view widely held and most prominently expressed by Sir Edward Coke, Hobbes’s main source on the Laws of England – it is only true if “reason” is understood as *per definitionem* consistent with deference to the sovereign.<sup>64</sup>

---

58 Thomas Hobbes, *Behemoth*, ed. F. Tönnies (London: Frank Cass, 1969), 51

59 Hobbes, *Behemoth*, 58

60 See Thomas Hobbes, *Dialogue between a Philosopher and a Student, of the Common Laws of England*, in: *Writings on Common Law and Hereditary Right*, ed. A. Cromartie and Q. Skinner (Oxford: Clarendon, 2008), 21

61 See Hobbes, *Dialogue*, 16

62 See Hobbes, *Dialogue*, 17

63 Hobbes, *Dialogue*, 13, emphasis added

64 For a new detailed study of the intellectual context and point of the Dialogue (which supports my use



It looks as though a new Hobbes is writing these lines. And in some sense, it is indeed a new Hobbes. However, this essay has shown that we do not have to be all too surprised. In his *Leviathan*, Hobbes is not just propounding considered philosophical judgements, he is also fighting internal philosophical struggles,<sup>65</sup> and if this reconstruction of his intellectual development is anything near correct, then Hobbes's aim of giving a convincing argument against the advisability of rebellion is much more constant and much more deeply rooted in Hobbes's oeuvre than are the premises and sub-premises employed *within* this project. His latest works can be seen as exploring quite new ways of inviting us to "enjoy [our] present state ... rather than go to war": no longer by arguing for the rationality of the making and keeping of certain contracts, no matter when and where, but by asking us to recognise the rationality inherent in the legislation of our actual commonwealth. This, I concede, is nothing less than a major turn. But if I am correct, then it is a turn for which an explanation – indeed a reason – is forthcoming.

Sometimes we have to see our heroes struggle in order to understand the nature of their battles.

---

of it) see the introduction by Alan Cromartie in the Clarendon edition.

65 Of course, Hobbes also had non-philosophical aims in his writing. However, I maintain that the theses defended in this paper are compatible with almost *any* non-philosophical diagnosis of his statements. For example, if the conciliatory tone of the *Dialogue* (see Cromartie's introduction, p. xix) is explained by recourse to Hobbes's need to get along with England's new rulers, there is still room for the thesis that there are also philosophical reasons of the kind outlined in this paper.