

# Hobbes and the Two Faces of Ethics

Arash Abizadeh



*Qua Iustitiae venerans Prudentia limen,  
Quid poterit vasto sanctus orbe coli.*



## HOBBS AND THE TWO FACES OF ETHICS

Reading Hobbes in light of both the history of ethics and the conceptual apparatus developed in recent work on normativity, this book challenges received interpretations of Hobbes and his historical significance. Arash Abizadeh uncovers the fundamental distinction underwriting Hobbes's ethics: between prudential reasons of the good, articulated via natural laws prescribing the means of self-preservation, and reasons of the right or justice, comprising contractual obligations for which we are accountable to others. He shows how Hobbes's distinction marks a watershed in the transition from the ancient Greek to the modern conception of ethics, and demonstrates the relevance of Hobbes's thought to current debates about normativity, reasons, and responsibility. His book will interest Hobbes scholars, historians of ethics, moral philosophers, and political theorists.

ARASH ABIZADEH is Associate Professor at the Department of Political Science and Associate Member of the Department of Philosophy at McGill University in Montreal. He has published extensively on Hobbes in journals including the *Journal of the History of Philosophy*, *The Historical Journal*, *Philosophers' Imprint*, and *Modern Intellectual History*.



# HOBBS AND THE TWO FACES OF ETHICS

ARASH ABIZADEH

*McGill University, Montreal*



**CAMBRIDGE**  
UNIVERSITY PRESS

**CAMBRIDGE**  
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom  
One Liberty Plaza, 20th Floor, New York, NY 10006, USA  
477 Williamstown Road, Port Melbourne, VIC 3207, Australia  
314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India  
79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.  
It furthers the University's mission by disseminating knowledge in the pursuit of  
education, learning, and research at the highest international levels of excellence.

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9781108417297](http://www.cambridge.org/9781108417297)

DOI: [10.1017/9781108277310](https://doi.org/10.1017/9781108277310)

© Arash Abizadeh 2018

This publication is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without the written  
permission of Cambridge University Press.

First published 2018

Printed and bound in Great Britain by Clays Ltd, Elcograf S.p.A.

*A catalogue record for this publication is available from the British Library.*

ISBN 978-1-108-41729-7 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of  
URLs for external or third-party internet websites referred to in this publication  
and does not guarantee that any content on such websites is, or will remain,  
accurate or appropriate.

# Contents

<i>Preface</i>	<i>page</i> vii
<i>List of Abbreviations</i>	ix
Introduction	I
Methodological Preliminaries	11
On Normativity and ‘Reason’	13
The Overarching Argument	17
PART I THE METAETHICS OF REASONS	25
1 Naturalism	27
1.1 An Error Theory	29
1.2 Descriptivist Reductionism	39
1.3 Reasoning-Based Descriptivist Reductionism	47
1.4 Noncognitivist Prescriptivism	52
1.5 A Hybrid Theory: Reasoning-Based and Prescriptivist	59
2 Mind, Action, and Reasoning	62
2.1 Practical Deliberation Is Partly Cognitive	65
2.2 Is Hobbesian Reasoning Passive?	72
2.3 Linguistic Reasoning is Reflectively Reason-Responsive	78
2.4 The Failure of the Hybrid Reading	90
PART II REASONS OF THE GOOD	95
3 Subjectivism, Instrumentalism, and Prudentialism about Reasons	97
3.1 Conative Subjectivism about Reasons? The Problem of Instrumental Transmission	99
3.2 Instrumentalism about Reasons? The Problem of Time	105
3.3 Prudentialism about Reasons	109

3.4	Cognitive Subjectivism about Reasons? Epistemically Relativized Objective Reasons	119
3.5	Suicide, Laws of Nature, and a Life Worth Living: Self-Preservation Is Not Survival	131
4	A Theory of the Good: Felicity by Anticipatory Pleasure	139
4.1	Four Distinct Questions about Goodness and 'Good'	141
4.2	The Meaning of 'Good': The Customary versus Reforming Sense	143
4.3	What Makes a Life Good: Anticipatory Pleasure	155
4.4	Two Complications for the Meaning of 'Good': Prescriptively Subversive Circumstances	164
4.5	Prescriptively Self-Fulfilling Circumstances	172
	<b>PART III REASONS OF THE RIGHT</b>	<b>181</b>
5	Accountability and Obligations	183
5.1	Three Types of Responsibility and Blame: Non-Normative, Critical, and Reactive	184
5.2	Directed Obligations	191
5.3	The Interest and the Will Theories of Direction	195
5.4	Owing an Obligation versus Being Held Accountable	199
5.5	Legal Accountability and Punishment	209
5.6	No Accountability for Intentions	212
6	The Laws of Nature, Morality, and Justice	217
6.1	The Meaning of 'Moral'	219
6.2	Reasons of the Right Cannot Be Derived from Reasons of the Good	223
6.3	The Relation between Natural Law and Obligation	228
6.4	The Relation between Natural Law and Civil Law	235
6.5	Two Puzzles about Natural Right and Natural Law	242
7	Rational Agency and Personhood	245
7.1	Entities, Voluntary Agents, and Rational Agents	247
7.2	Persons: Representer or Represented?	249
7.3	Natural Persons Are Authors	252
7.4	Artificial Persons: True and By Fiction	255
7.5	Accountability Is Interpersonal: No Accountability to Oneself	258
	Conclusion: Naturalism and Normativity	263
	<i>Works Cited</i>	277
	<i>Index</i>	286

## *Preface*

I didn't set out to write this book, or any other book on Hobbes. I thought I was writing a book on Rousseau. I had introduced Rousseau's account of language with a couple of paragraphs on Hobbes. But they were never quite right, so they kept expanding until too bloated to stay in a book on Rousseau. Having extracted the offending paragraphs – they were to compose a stand-alone article – yet still unsatisfied, I decided to teach a seminar on Hobbes and Descartes to figure it out. That was 2006. The seminar inspired two more articles on Hobbes, and by 2007, while on sabbatical, I discovered a book on Hobbes was being written. By 2008, it had been drafted. But I was still unsatisfied, so it kept expanding until there were too many words to stay together. I split the book, in 2011, into two. The first had two chapters needing revisions. The first chapter had two problems; I resolved to write one article on each problem, after which I'd revise the chapter with solution in hand. So near the end of 2013, I set out to write the first of two articles, about the first of two problems with the first of two chapters, in the first of two draft manuscripts on Hobbes. This book is that article.

This book is my first, so I begin with thanks to my teachers: Richard Noble, who inspired my love for this stuff; Jerry Cohen, who taught me how to make an argument; David Miller, who suggested I find inspiration in what I disagree with; Seyla Benhabib, who urged me to link past to present; Bonnie Honig, who prompted me to peek past authorial intentions; Pratap Mehta, who dazzled me with graceful erudition; and Richard Tuck, who engaged me like a colleague from the start and, of course, taught me Hobbes.

This monograph benefitted tremendously from two manuscript workshops, at McGill in 2015 and QMUL in 2016. Many thanks to the participants, particularly those who gave formal commentaries: Signey Gutnick Allen, Adrian Blau, Terence Cuneo, Robin Douglass, Evan Fox-Decent, Kinch Hoekstra, Travis Smith, Susanne Sreedhar, Sarah Stroud,

Laurens van Apeldoorn, and especially Stephen Darwall, whose intellectual generosity, acumen, and range are an inspiration. For comments I also thank Libby Barringer, Deborah Baumgold, Jérémie Duhamel, Julia Driver, Pablo Gilabert, Ian MacMullen, Al Martinich, Christian Nadeau, Kieran Oberman, Davide Panagia, Martine Pécharman, Paul Sagar, Tim Stanton, Mónica Brito Vieira, and audiences at Berkeley, Edinburgh, Montréal, Princeton, UCLA, WUSL, and York. I must also thank Lucas Stanczyk, whose conversation incited me to write this book. Thanks above all to Johan Olsthoorn, one of Cambridge University Press's referees, whose extensive criticisms and suggestions were frankly amazing.

I drafted the bulk of this book in autumn 2014, during a half-sabbatical at the Centre de Recherche en Éthique at the Université de Montréal. I am grateful to my colleagues there for their hospitality. Thanks are also due to my McGill political theory colleagues, who have made my intellectual home so stimulating: Jacob Levy, Catherine Lu, Victor Muñiz-Fraticelli, Will Roberts, Hasana Sharp, Daniel Weinstock, and Yves Winter.

Thanks to my editor, Hilary Gaskin, for supporting this project, and SSHRC for financial support. Parts of **Chapter 2** first appeared in *Journal of the History of Philosophy* 55.1 (2017): 1–34 (Copyright © 2017 The Johns Hopkins University Press). Parts of **Chapter 7** first appeared in *The Historical Journal* 60.4 (2017): 915–941 (Copyright © 2017 Cambridge University Press).

Hobbes said fear and he were born twins together. Isaiah, Esmée, and Dante were born with Hobbes their companion: they have never had a father who wasn't writing a book on Hobbes. I am grateful to them for the combination of sleep deprivation, curiosity, and wonder that has indelibly marked this book. My spouse Noelle, more than anyone, is thankful it is done. I dedicate this book to her.

## Abbreviations

### Hobbes's Works

- EW *The English Works of Thomas Hobbes of Malmesbury*, ed. William Molesworth. 11 vols. London: John Bohn, 1839–1845.
- OL *Opera Philosophica Quae Latine Scripsit Omnia*, ed. William Molesworth. 5 vols. London: John Bohn, 1839–1845.
- EL *The Elements of Law*. Harley MS 4235 (British Library, London) [chapters renumbered consecutively]. [1640]
- O *Objectiones to Descartes*. [*The Philosophical Writings of Descartes*, ed. J. Cottingham, R. Stoothoff, and D. Murdoch (Cambridge: Cambridge University Press, 1984).] [1641]
- DCv *De Cive: The Latin Version*, ed. Howard Warrender (Oxford: Clarendon Press, 1983). [*De Cive: The English Version*, ed. Howard Warrender (Oxford: Clarendon Press, 1983).] [1642, 2nd ed. 1647]
- AW *Critique du De Mundo de Thomas White*, ed. J. Jacquot and H. W. Jones (Paris: Vrin-CNRS, 1973). [*Thomas White's De Mundo Examined*, trans. H. W. Jones (London: Bradford University Press, 1976).] [1642–1643]
- LN *Of Liberty and Necessity, in Hobbes and Bramhall on Liberty and Necessity*, ed. Vere Chappell (Cambridge: Cambridge University Press, 1999). [1645]
- MDO *A Minute or first Draught of the Optiques*. Harley MS 3360 (British Library, London). [1646]
- L and LL *Leviathan: The English and Latin Texts*, ed. Noel Malcolm (Oxford: Clarendon Press, 2012). [1651, 1668]
- DC *De Corpore*, in OL 1 [EW 1]. [1655]

- Q *The Questions Concerning Liberty, Necessity, And Chance* (London: Andrew Crook, 1656). Pages numbers after ‘/’ are to EW 5. [1656]
- DH *De Homine*, in OL 2. [*Man and Citizen (De Homine and De Cive)*, ed. Bernard Gert (Indianapolis, IN: Hackett Publishing, 1991).] [1658]
- AB *An Answer to a Book Published by Dr. Bramhall, late Bishop of Derry; called The Catching of Leviathan. Together With an Historical Narration Concerning Heresie, And the Punishment thereof* (London: For W. Crooke at the Green Dragon, 1682). Page numbers after ‘/’ are to EW 4. [1668]
- HNH *An Historical Narration Concerning Heresie, And the Punishment thereof* (London, 1680). Page numbers after ‘/’ are to EW 4. [1668]
- B *Behemoth, or, The Long Parliament*, ed. Paul Seaward (Oxford: Clarendon Press, 2010). [c. 1669]
- HE *Historia Ecclesiastica*, ed. P. Springborg, P. Stablein, and P. Wilson (Paris: Honoré Champion, 2008). [c. 1671]
- D *A Dialogue Between a Philosopher and a Student, of the Common Laws of England, in Writings of Common Law and Hereditary Right*, ed. A. Cromartie and Q. Skinner (Oxford: Clarendon Press, 2005). [1681]

### Other Primary Texts

- Laws Plato. *The Laws*. [c. 348 BCE]
- Rh Aristotle. *On Rhetoric*. [c. 340–335 BCE]
- NE Aristotle. *Nicomachean Ethics*. [c. 330 BCE]
- Ex Epicurus. *Epicurus: The Extant Remains*, ed. C. Bailey. Oxford: Clarendon Press, 1926. [lived 341–270 BCE]
- DRN Lucretius. *De Rerum Natura*. [c. 75–54 BCE]
- DO Cicero. *De Officiis*. [44 BCE]
- DF Cicero. *De Finibus Bonorum et Malorum*. [45 BCE]
- DI Seneca. *De Ira*. [41–52 CE]
- Lives Diogenes Laertius. *Lives of Eminent Philosophers*. [200s CE]
- ST Thomas Aquinas. *Summa Theologicae*. [c. 1265–1274]
- DLDL Francisco Suárez. *De Legibus ac Deo Legislatore*. 1612.
- JBP Hugo Grotius. *De Jure Belli ac Pacis*. 1625. 2nd ed. 1646.

- DTL John Bramhall. *A Defence of True Liberty from Antecedent and Extrinsecall Necessity*. 1655.
- BL John Aubrey. *Brief Lives*. [Ed. A. Clark. 2 vols. Oxford: Clarendon Press, 1898.] [1669–1696]
- THN David Hume. *A Treatise of Human Nature*. 1739–1740.
- SLRI John Stuart Mill. *A System of Logic Ratiocinative and Inductive*. 1843.



## Introduction

There is no justice, or if there is, it is supreme folly, because attending to the advantage of others is violence to one's own.

– Grotius's "Carneades"<sup>1</sup>

The seventeenth century is a watershed in the history of European ethics – a moment in which a eudaimonistic model rooted in ancient Greece began to give way to a distinctly modern, juridical model of morality. If the central question of Greek ethics was how one should live, the answer lay in the nature of the ultimate good. Aristotle observed that everyone agreed in calling the ultimate good of a human life *eudaimonia* or well-being; disagreement was over its constituents.<sup>2</sup> On the eudaimonistic view, an account of the good life and how to achieve it covers the whole of ethics, because all of one's practical reasons for action and affective reasons for desire or passion are rooted in one's own good: eudaimonism is an ultimately egoistic ethics.

The modern conception of ethics departed from this picture in two significant ways. First, if Greek ethics specified the dispositions of character needed to realize eudaimonia, modern ethics took the form of a juridical *code*, i.e., moral laws and obligations. Second, such obligations may in principle conflict with one's own good.<sup>3</sup> True, the basis for this shift had been laid much earlier – by the Stoics and Cicero, who introduced the notion of natural law, and by Aquinas, who fused natural law with the Christian idea of a divine legislator. But the decisive break occurred in the seventeenth century: notwithstanding the earlier legalistic framework, seventeenth-century ethics is distinguished by the emergence, through the works of Francisco Suárez, Hugo Grotius, and Thomas Hobbes, of a *juridical* notion of obligation.

Suárez fired an opening salvo in this direction from the apex of late scholasticism in his 1612 *De Legibus ac Deo Legislatore*. He did so in the

<sup>1</sup> JBP Prolegomena.5.    <sup>2</sup> NE I.4.    <sup>3</sup> Sidgwick (1896); Anscombe (1958); Darwall (1995, 2012).

course of seeking middle ground between intellectualists, who thought natural laws merely *indicate* what is intrinsically good by nature, and voluntarists, who insisted that God's will *makes* things good. Suárez argued that some things are indeed good by nature, but that law in its proper sense does not consist in merely pointing out what is good. Natural law therefore has two aspects. As a dictate of reason, it indicates what is necessary for virtue or honourableness (*honestas*) and hence for *felicitas* (as eudaimonia was rendered into Latin). But it is properly law, and hence obligatory, only insofar as it is *prescribed* or commanded by God. It is true that classical natural-law theorists such as Aquinas had also previously claimed that natural laws obligate, and hence are morally due or *debitum*. But they worked with a wholly eudaimonistic notion of obligation. For Aquinas, "obligatory" just means *necessary* to felicity – whether in the loose sense of conducive (*utile, melius, expediens*) to supererogatory virtue (yielding an admonition of *counsel*), or strictly necessary because indispensable to virtue as such (yielding a true debt and precept of *obligation*).<sup>4</sup>

Suárez answered Aquinas's twofold distinction with a threefold one. First, what is due because supererogatory or optimum, but which is not strictly necessary to felicity, is a matter not of "moral obligation," but counsel. Second, a dictate of reason merely indicating what is indispensable for felicity does not impose obligation either: it indicates a natural debt to oneself.<sup>5</sup> Obligation arises solely in virtue of a prescription binding on pain of guilt, and this requires not counsel, but the *command* of a superior to whom one is accountable for violations.<sup>6</sup> God's command *superimposes* obligation on the "natural" *honestas* of what is necessary for felicity – it cannot be reduced to eudaimonistic necessity.<sup>7</sup> Two points are noteworthy. First, obligation is inherently grounded for Suárez in God's will. Second, despite his juridical conception of obligation, Suárez retained an essentially eudaimonistic outlook: while the law's obligatory character renders one *accountable* to a superior, one's *reason* for abiding by the law is not that it is obligatory, but that it specifies the path to felicity, i.e., the reason is extrinsic to the obligation. Grotius broke with Suárez on both counts.

Grotius's response to the classical objection against justice is emblematic of this watershed. Grotius raised the objection in Carneades's voice (quoted in our epigraph) at the start of his prolegomena to *De Jure Belli ac Pacis* (1625, second edition 1631). Grotius had Carneades argue that there

<sup>4</sup> ST I-II.99.1; 99.5; 100.2; 108.4; II-II.58.3; 88.3.     <sup>5</sup> DLDL II.9.7; 7.11–12; 6.11; 9.1; 9.4.

<sup>6</sup> DLDL I.1.7–2.9; 3.18–4.4; 12.4.     <sup>7</sup> DLDL II.6.11–12.

is no *jus naturale*, no natural norms of justice: “men imposed laws [*jura*] upon themselves,” purely for reasons of mutual expedience (*utilitate*); these conventions consequently lose their normative purchase the moment they turn to one’s disadvantage. The classical eudaimonistic response to this sort of objection began by granting its basic premise: practical reasons are all ultimately rooted in one’s own good. But apart from Epicureans, the eudaimonists typically held that the virtues – including justice, which requires attending to a common good – are themselves intrinsic components of one’s own good. Cicero argued on this basis, for example, that because what is expedient or *utile* depends on what constitutes one’s good, no action is expedient unless it is *honestus*.<sup>8</sup> Hence the classical response was that the *reason* one should care for others is that doing so is a constituent of – and not merely instrumental to – one’s own good. This assumption of an underlying natural harmony of interests is why classical natural-law theory remained eudaimonistic. To be sure, natural law, as Aquinas and Suárez defined it, is oriented to the *common* good, but it is *normative* for individuals because it intrinsically directs each to their own good: the common good is constitutive of each individual’s good.

Grotius’s response to Carneades in the prolegomena can be read as continuous with this eudaimonistic tradition.<sup>9</sup> There he argued that justice is not folly because, although sometimes it requires forgoing what is expedient in a narrowly self-regarding sense (*sibi utilia*), we are also naturally *sociable* creatures who long for living peaceably together – and the point of justice or *jus naturale* is to secure this end. Justice is not merely *instrumental* to sociable, peaceable living; living justly is also a way of *expressing* our sociable nature: “Even if no *utilitas* were expected from observing *juris*, it would yet be wisdom, not folly, to obey the felt direction of our own nature.”<sup>10</sup>

But once Grotius carried his response past the prolegomena, he entered new territory. There he distinguished *jus* understood expansively to mean whatever is not unjust and hence repugnant to a society of reasonable creatures – i.e., what is consistent with right reason – from *jus* meaning a *lex* or law consisting in “a rule of moral actions imposing obligation to what is right.”<sup>11</sup> The break with eudaimonism occurs with this notion of obligation. Like Suárez, Grotius insisted it is an “abuse” to include under obligation supererogatory acts that are “by nature *honestum*” and praiseworthy, “but not truly due [*debitum*].”<sup>12</sup> To characterize natural law, “we

<sup>8</sup> DO 2.10; 3.12; 3.116–17. <sup>9</sup> Irwin (2008). <sup>10</sup> JBP Prolegomena.6–18; 44. <sup>11</sup> JBP I.1.9.  
<sup>12</sup> JBP II.14.6; 5.9.

require obligation: for counsels and other such prescriptions, which are *honestas* but not obligatory, do not fall under the name *legis* or *juris*.<sup>13</sup> He then distinguished a loose and a strict sense in which acts can be truly due or obligatory under *jus naturale*. The loose sense of obligation – which corresponds to Aquinas’s strict notion – is eudaimonistic: it “signifies what cannot be omitted without dishonour [*inhoneste*].”<sup>14</sup> But the strict notion – “obligation imposed by expletive justice”<sup>15</sup> – is a juridical one, which Grotius modelled on the Roman law of actions.<sup>16</sup>

To have an *actio* in Roman law was to have available a lawful procedure – either judicial or private – to vindicate violations of one’s claims.<sup>17</sup> Similarly, the essence of being under an obligation in Grotius’s juridical sense is for someone to have the standing *to hold one accountable* to its terms: to demand its fulfilment and seek its enforcement and, in case of violation, to seek reparations and punishment for injury.<sup>18</sup> This standing or *jus exigendi* is constitutive of what Grotius famously called a “perfect right.” To have an *imperfect* right is to merit something – to be worthy, apt, or fit for it. If you merit my gratitude (or charity or generosity), then my gratitude is due or obligatory in the loose sense that it is indispensable to my *honestas* – not only is it praiseworthy, but its omission would be blameworthy (*culpa*) and sinful (*peccatum*). But you have no standing to demand or enforce my gratitude – whether in court or by arms – and I am not bound to restitution.<sup>19</sup> A *perfect* right, by contrast, is due strictly: the correlative obligations are a matter of justice in the strict, “expletive” or reparative sense, and the right-holder has standing to demand and seek to enforce it.<sup>20</sup> To have an obligation here means to be liable for reparations and punishment in case of guilt (*culpa*).<sup>21</sup>

Grotius thus departed from Suárez on two counts. First, juridical obligation does not depend on God’s will: the standing to hold accountable is grounded directly in humanity’s sociable nature. Second, juridical obligation not only consists in the perfect-right-holder’s standing to hold accountable; it is also intrinsically normative for the obligated person: natural sociability simultaneously grounds both the standing to hold

<sup>13</sup> JBP I.1.9.    <sup>14</sup> JBP II.7.2.    <sup>15</sup> JBP II.7.2; 7.4; 13.1; 20.2.    <sup>16</sup> JBP I.1.5; II.1.2; 14.6.

<sup>17</sup> Metzger (1997).

<sup>18</sup> In war, these correspond to the rights of self-defence, recuperation, and punishment. JBP II.1.2.

<sup>19</sup> JBP II.11.3; 11.4; 13.17; 17.2; 17.9; 25.3; 22.6.    <sup>20</sup> JBP I.1.5; 1.8; II.7.9.

<sup>21</sup> Aquinas also held a notion of “legal debt” – which he distinguished from (both supererogatory and indispensable) moral debt – for which one is liable for reparations and punishment. ST II-II.80.1; 108.2. But for Aquinas, reparations and punishment arise from a general imperative to restore balance in a cosmic order upset by sin, not from someone’s particular standing to hold one accountable. I-II.87.

accountable and the obligated person's reason to fulfil the obligation.<sup>22</sup> This implies rational agents have two distinct types of practical reasons: reasons corresponding to eudaimonistic obligations grounded in one's own good; and reasons grounded in juridical obligation or *jus*. In principle these reasons could diverge, but in practice Grotius obviated this possibility: since sociability is not only constitutive of each individual's good but is also the grounds for juridical obligation, fulfilling one's obligations is constitutive of one's own good.

Like Grotius, Hobbes developed an intrinsically normative, juridical notion of obligation, even as he retained the traditional, eudaimonistic one. But he also took the radical step of severing juridical obligation from natural law and all obligation from natural sociability – thereby transforming the notions of obligation and natural law and their mutual relation. Unlike Grotius, Hobbes asserted that a person's juridical obligations, for which he is accountable to others, always arise “from some Act of his own,”<sup>23</sup> i.e., via conventions such as contract whereby a person signifies the intention to bind himself to others. Obligation in the proper, juridical sense is grounded neither in God's will (Suárez) nor in natural sociability (Grotius), but in the interpersonal meaning of voluntary acts. The pre-conventional laws of nature, by contrast, impose obligations only in the loose, eudaimonistic sense harking back to Aquinas: they prescribe the means to one's own good. Natural obligation is not obligation in the proper sense, just as natural law is not, according to Hobbes, law in the proper sense either – unless and until acknowledged conventionally as authoritative command.<sup>24</sup>

The upshot is that Hobbes's ethics comprises two distinct dimensions of normativity. The first comprises *reasons of the good*: reasons we might consider when reasoning from a first-personal perspective, and which observers might take us to have from the third-personal perspective, but for which we are not accountable to anyone. When we have normative reasons of this first kind, we are responsible for the passions or actions for which they are reasons, responsible in the sense that the passions or actions are *attributable* to us: we may be correctly counselled or warned, and justifiably commended or criticized, in their light. The attribution, counsel, and appraisal presume the rational capacity reflectively to understand the advice and appraisal, respond to the reasons involved, be guided by them, and, indeed, justify our passions or actions in their light. But they do not presume anyone has any claim to such justificatory responses.

<sup>22</sup> Darwall (2012). <sup>23</sup> L II.7: 154. <sup>24</sup> L 26.8: 418; 15.41: 242.

The second, distinct dimension comprises *reasons of the right*: reasons for which we are second-personably *accountable* to others. The reasons themselves and others' standing to hold us accountable for them are grounded in the interpersonally recognized signs of our will. Others have standing to demand our conformity and to react to failures to conform to such reasons in ways that reiterate and seek to vindicate the demand: to condemn and sanction failures by demanding excuses, justification, or acknowledgement of a wrong and hence apology, compensation, or redress. They have an auxiliary claim to our normative response, which claim presumes the capacity to recognize relationships with others and the demands that constitute them as normative. Failing to heed the first kind of reason renders us an appropriate target for criticism or *critical* blame, but failing to heed the second kind renders us an appropriate target for vindicatory or *reactive* blame.<sup>25</sup>

There is, moreover, a fundamental chasm between the attributability and accountability dimensions of Hobbes's ethics: reasons of the right are neither reducible to nor wholly derivable from reasons of the good. The attributability dimension paradigmatically consists in the rational precepts of natural law, the most important type of which prescribe to each the social means of self-preservation; the accountability dimension paradigmatically comprises the obligations arising from contract, for which one is accountable to others. Hobbes signalled the chasm between these two dimensions in the English *Leviathan's* (1651) table of sciences, where he distinguished "ETHIQUES," which concerns "Consequences from the *Passions* of Men," from "The *Science* of JUST and UNJUST," which concerns "Consequences from *Speech*."<sup>26</sup> "Ethiques" is restricted to the traditional, eudaimonistic dimension of normativity grounded in an agent's own good; and "Moral Philosophy" – which "is nothing but the Science of what is *Good*, and *Evill*, in the conversation, and Society of man-kind"<sup>27</sup> – is that part of ethiques dealing specifically with social *relations*. But whereas the science of the good concerns *natural* laws and eudaimonistic obligations, the science of justice concerns *artificial* laws and juridical obligations. Fulfilling contractual obligations is also prescribed by natural law, of course, but this is because in Hobbes's view we have prudential reasons to heed reasons of the right – not because natural law furnishes or grounds reasons of the right.

That justice coincides with prudence in this way is precisely what the "Foole" in *Leviathan* denies. The Foole echoes the objection raised by

<sup>25</sup> On the attributability and accountability senses of responsibility, see Watson (1996).

<sup>26</sup> L 9: 131.    <sup>27</sup> L 15.40: 242.

Grotius's "Carneades,"<sup>28</sup> but the classical, eudaimonistic response canvassed by Grotius was unavailable to Hobbes. For in repudiating natural sociability,<sup>29</sup> Hobbes was denying not only that juridical obligations derive from or reduce to natural ones but also that the common good is constitutive of one's own good. If a law of nature prescribes what is in the common good, it does so only because it is an *instrument* to the individual's own good: the eudaimonistic obligations it imposes are not intrinsically oriented to the common good. Once the classical assumption of an underlying harmony of interests is repudiated, one faces a potentially fundamental conflict between one's own good and the good of others; and once eudaimonistic and juridical obligations are given distinct grounds, one faces a potential conflict between reasons of the good and of the right. Hobbes was pivotal amongst early modern thinkers for taking these two possibilities so seriously. As I hope to show, he fought the first danger via a common good – peace – constructed from the overlap between each individual's antecedent good; and he fought the second danger via his sign theory of consent, through which he built prudential constraints into the content of juridical obligations. He had to manage this potential gap, between reasons of the good and of the right, because on his account one can be moved only by what appears to be good.

The seventeenth century was pivotal in the history of not only normative ethics, but also metaethics. This is the century in which a mechanistic model of science began to displace the older, teleological model, and Hobbes was at the forefront of this development, alongside thinkers such as Isaac Newton, Galileo Galilei, Pierre Gassendi, Marin Mersenne, and René Descartes. Hobbes was especially concerned with the implications of the new model for ethics. Any comprehensive treatment of Hobbes's ethics therefore faces the challenge of how to reconcile his extensive normative-sounding language with his uncompromisingly mechanistic metaphysics and materialist account of language. On the one hand, Hobbes suggested that the only real entities are extended bodies in motion, all of whose real properties can be reductively analyzed and redescribed in terms of extension and motion;<sup>30</sup> he also claimed positive names are meaningful only if they are names of bodies, their properties, conceptions of them, or linguistic expressions.<sup>31</sup> Hobbes even went so far as to produce a materialist account of mathematical objects, according to which a point, for example,

<sup>28</sup> Tuck (1987). <sup>29</sup> DCv 1.2. <sup>30</sup> LL 9: 125; AW 3.2.

<sup>31</sup> L 4.14–20: 58–60; DC 5.2. "Negative" names (L 4.19: 60) and words that are not names, such as the copula 'is' (L 46.16: 1078), can also be meaningful.

is an extended body whose magnitude is simply not considered (or is considered to be zero) for purposes of demonstration.<sup>32</sup> On the other hand, Hobbes deployed a vast array of normative vocabulary – ‘good’, ‘natural law’, ‘duty’, ‘obligation’, ‘ought’, ‘right’, ‘justice’, and ‘reasons’ – that seem to attribute properties to things and events not reducible to extension and motion.

There have been two broad approaches to this puzzle. One strategy, which takes Hobbes’s normative vocabulary as evidence for a rich and genuinely normative philosophy, has been to detach the normative branch of his philosophy from his natural philosophy, and to ground the former in God’s will.<sup>33</sup> This “theological” approach has the merit of taking Hobbes’s normative vocabulary seriously, but runs against his claim to have built a unified, scientific-philosophical system. The problem is not merely that these two branches of Hobbes’s science or philosophy<sup>34</sup> are detached from and rendered incompatible with each other, but that his ethics and political philosophy are rendered incompatible with the very conception of science upon which he insisted: science, for Hobbes, restricts itself to propositions conceivable to the human mind, and therefore excludes theology and appeal to God.<sup>35</sup>

The second approach has sought to preserve the link between the two branches, and to honour Hobbes’s aspiration to develop a science of morals and politics, by attributing an ethical theory to him that deflates his normative claims. Thus Hobbes has frequently been read as a subjectivist about reasons and value<sup>36</sup> and, at the metaethical level, as an ethical-naturalist reductionist – according to whom all normative properties and facts reduce to non-normative natural properties and facts (concerning, for example, the means for fulfilling one’s own desires) – or as an error theorist or noncognitivist who simply denied the existence of normative properties. None of these readings, I shall argue, are plausible: Hobbes was committed to the view that some facts provide irreducibly normative reasons to believe, desire, or act, and he was not a subjectivist about reasons or value. Indeed, Hobbes posited two distinct dimensions of normativity. Hobbes’s implicit, broadly naturalist metaethics is best understood as denying that normative properties are *real* properties even while affirming

<sup>32</sup> Jessep (1999: 76–80).

<sup>33</sup> Taylor (1965); Warrender (1957). Strauss (1963) also detaches Hobbes’s normative from his natural philosophy.

<sup>34</sup> Seventeenth-century writers used these two terms synonymously.

<sup>35</sup> DC 1.8–9. On Hobbes’s conception of science, see Jessep (1996, 2010).

<sup>36</sup> Nagel (1959); Gauthier (1969); Kemp (1970); Watkins (1973); Hampton (1986).

that irreducibly normative propositions are truth apt and indeed sometimes true.

These claims about Hobbes's ethics are significant for several reasons. First, the distinction between the two dimensions of normativity is essential to resolving an apparent inconsistency plaguing two of Hobbes's most central concepts: the right of nature and the law of nature. Hobbes asserted that in the state of nature, prior to any conventionally incurred obligations, rational agents each possess a right of nature comprising a liberty-right to do whatever they judge to be relevant means for self-preservation (where a liberty-right implies the absence of obligation to do or forbear).<sup>37</sup> He then subsequently claimed this is equivalent to a liberty to do anything at all.<sup>38</sup> One puzzle is how these two characterizations could be equivalent: the former seems conditional on the individual's subjective judgement; the latter does not.<sup>39</sup> A second puzzle arises from the fact that Hobbes also asserted that rational agents in the state of nature are already obliged by the laws of nature. Yet if rational agents are always already "obliged" by natural law, how can they also be entirely free from obligation, as the right of nature implies? The apparent contradiction lies right there on the surface; Hobbes explicitly announced it in *Leviathan*, asserting that "RIGHT, consisteth in liberty to do, or to forbear; Whereas LAW, determineth, and bindeth to one of them: so that Law, and Right, differ as much, as Obligation, and Liberty; which in one and the same matter are inconsistent."<sup>40</sup> The answer to both puzzles lies in the fact that the right of nature concerns the absence of obligations in Hobbes's proper, juridical sense for which one is *accountable* to others – reasons of the right – whereas the laws of nature intrinsically "oblige" only in the loose, eudaimonistic sense that they articulate reasons governing passions and actions that are *attributable* to one but for which one is not intrinsically accountable – reasons of the good.

Second, the attributability and accountability dimensions of normativity and responsibility are reflected in Hobbes's distinction between rational agency and personhood: while any rational agent may have, be attribution-responsible for, recognize, respond to, and be guided by reasons, only persons can be accountable to others. Hobbes's celebrated notion of

<sup>37</sup> I use 'claim-right', 'liberty-right', 'immunity-right', and 'power-right' in their Hohfeldian senses (Hohfeld 1964). Yates (2013) provides a defence of using the Hohfeldian apparatus for interpreting Hobbes.

<sup>38</sup> For the two formulations, see L 14.1: 198 and 14.4: 198; for liberty-right as the absence of obligation, see 14.3: 198.

<sup>39</sup> Kavka (1986: 300–302). <sup>40</sup> L 14.3: 198.

personhood is, even in the case of “natural persons,” an intrinsically artificial construct whose full significance is appreciated only in light of this distinction; the potential for accountability is its distinctive normative feature.

Third, almost all interpretations of Hobbes on offer deny any fundamental chasm between these two dimensions. On the theological interpretation, the obligation to fulfil voluntary contracts is grounded in the prior obligation to obey God who, via the third law of nature, commands their fulfilment.<sup>41</sup> On the “orthodox” interpretation, contracts “obligate” only in the sense that one has an instrumental or prudential reason to fulfil their terms.<sup>42</sup> Still other interpreters, according to whom genuine obligations arise only with political society and positive law, argue that the normative force of positive law derives from natural law.<sup>43</sup> On all these interpretations, there is no genuine chasm between the laws of nature and conventionally incurred obligations: the obligation to fulfil contracts is entirely grounded in natural law.<sup>44</sup> Moreover, if, as the orthodox interpretation suggests, all reasons for action derive from reasons for taking the relevant means for fulfilling one’s own desires, then Hobbes’s philosophy seems inhospitable to genuinely *moral* reasons in any modern sense – a conclusion drawn by many of Hobbes’s readers.<sup>45</sup> Focussing on the accountability dimension of normativity helps clarify where these readings go wrong: insofar as they imply there are no genuine reasons of the right, they fail to account for an essential feature of Hobbes’s philosophy. Such a focus helps us discern not only Hobbes’s commitment to irreducibly normative facts but also the sense in which he provided what we would now recognize as a genuinely moral philosophy.

Fourth, clarifying Hobbes’s implicit metaethics forces us to rethink his role in the history of metaethical naturalism: far from being seen as a founder of reductionism, he must be recognized as a forerunner of those seeking to reconcile their metaphysical naturalism with irreducible normativity.

Finally, the present work exposes Hobbes’s pivotal role in setting the stage for a distinctly modern conception of morality. On the classical natural-law view, specifically *moral* reasons are those oriented to the

<sup>41</sup> Warrender (1957).

<sup>42</sup> Nagel (1959); Gauthier (1979: 557); Skinner (2002: chapter 9). Cf. Darwall (1995).

<sup>43</sup> Bobbio (1993).

<sup>44</sup> Hoekstra’s (2003: 113) assertion that “the duty to obey the sovereign one has covenanted to obey depends on a prior duty to obey the law of nature” is a commonplace of Hobbes scholarship.

<sup>45</sup> E.g. Nagel (1959). For Hobbes’s contemporaries, see Mintz (1969); Parkin (2007).

common good: this orientation is why the eudaimonistic obligations imposed by natural law are moral obligations. Moral reasons are here viewed as a species of eudaimonistic ones. But on the rival, juridical view, moral reasons are a distinct genus: they consist in reasons for which others have standing to hold one accountable. Unlike Suárez and Grotius, Hobbes fully abandoned the classical model of a basic harmony of interests. And he was – despite his thoroughly prudentialist reputation – one of the pivotal founders of the juridical model that would flourish in the subsequent century.

### Methodological Preliminaries

Recent work on normativity has considerably advanced our understanding of the differences between distinct concepts that traditionally in English have been expressed interchangeably via terms such as ‘reason’, ‘reasons’, ‘reasoning’, and ‘rational’. Since in this book I deploy some of these distinctions to clarify Hobbes’s normative philosophy, I here wish to defend the appropriateness of this deployment.

The interpretation of historical philosophical texts is inherently fraught with the risk of anachronistic distortion. The question is whether bringing contemporary philosophical language to bear on seventeenth-century philosophers’ writings ends up distorting their philosophy more than illuminating it. On the one hand, if Hobbes was implicitly making conceptual distinctions that, for want of a technical vocabulary, he did not make explicit, then making those distinctions explicit by expressing his philosophy in contemporary vocabulary can clarify his thought. (This is the case, I will show, for the senses of ‘reason’ philosophers now carefully distinguish.) I would go so far as to claim that even if Hobbes himself did not, even implicitly, make distinctions we now deem relevant, articulating his philosophy in their light can tell us something important about his philosophical system.<sup>46</sup> (An example I discuss in [Chapter 4](#) is the distinction between hedonist and desire-fulfilment theories of the good.)

On the other hand, commentators who help themselves to contemporary tools must protect the scholarly integrity of their interpretation by asking whether the contemporary *vocabulary* on which they draw would have made sense to Hobbes himself had he been confronted with it, or whether instead it presupposes a *conceptual* apparatus that would have been unintelligible to him and his contemporaries. If the conceptual apparatus

<sup>46</sup> Blau (2017).

would have been *unintelligible*, then deploying the vocabulary expressing it to interpret his philosophy cannot but distort our understanding of it. But intelligibility is an extremely low standard to meet. By contrast, the conceptual apparatus may very well have been *unavailable* in his period, in which case interpreters carry the burden of showing that our contemporary vocabulary clarifies his expressed philosophy. Discharging any such a burden requires satisfying an interpretive principle of charity.<sup>47</sup> Interpreters may even seek to show that Hobbes, despite lacking the vocabulary fully to articulate it, was engaged in the project of forging a new conceptual apparatus to *make* it available to his readers, and that our contemporary vocabulary more efficiently expresses this new apparatus than the one available at the time.<sup>48</sup> This possibility should hardly be surprising for a figure of Hobbes's stature who, in one historian's judgement, "created English-language philosophy."<sup>49</sup> Indeed, this is precisely what I shall show Hobbes was doing in forging a distinction between two dimensions of normativity: the vocabulary and conceptual apparatus for this distinction were barely available prior to Hobbes, and Hobbes himself was not always explicit about the distinction. Nor was he even always fully cognizant of its implications; the significance, novelty, and indeed content of Hobbes's philosophy are not reducible to mental states in his head.

A principle of charity does not rule out inconsistencies in Hobbes's oeuvre: one must be alive to the possibility that Hobbes reasoned erroneously, failed to appreciate the consequences of a particular position, or changed his mind. But confronted with an interpretation that, by deploying contemporary vocabulary, illuminates Hobbes's texts without imputing inconsistency to him, rival interpretations according to which Hobbes made mistakes or changed his mind require a concrete argument. By contrast, if contemporary vocabulary creates more puzzles than it solves, then, even though it might still serve the critical function of *appraising* a philosopher's thought, it cannot serve as the vocabulary used to express it. Sometimes this burden is not met, and I shall give important examples in relation to the implicit metaethical commitments underlying Hobbes's view: ethical-naturalist reductionism, the error theory, and prescriptivism are our concepts, not Hobbes's, and *imputing* them to him requires that they clarify more than obscure his philosophy. If the claim is the weaker one – that although the concept in question was unavailable to him, his *philosophy* is nevertheless fruitfully read as an instantiation of it – then the evidence for imputing it must be indirect: one must show his commitment

<sup>47</sup> Davidson (1984: chapter 13).    <sup>48</sup> Pace Skinner (1988: 77–78).    <sup>49</sup> Tuck (1989: vii).

to substantive positions entailing propositions definitive of our contemporary concept. Imputing these particular concepts to Hobbes, however, makes a mess of his philosophy: they are irreconcilable with too many of his commitments. Someone might acknowledge this counter-evidence, but insist that Hobbes *aspired* but failed to deliver a reductionist, error-theoretical, or prescriptivist metaethics.<sup>50</sup> But this is a risky interpretive strategy, since again the most relevant evidence in favour would be the fact that Hobbes *did* live up to their requirements. To impute to Hobbes a concept unavailable to him, in order then to indict him for failing to live up to its requirements, is fully to fall prey to the risk of anachronism: it is on balance to distort the structure of his normative philosophy.

My hope is that the interpretation offered in this book draws on contemporary tools in ways that illuminate more than obscure Hobbes's philosophy and its historical significance. The following section's purpose is to vindicate this hope by mapping out how Hobbes's own reason-talk corresponds to many of the technical distinctions explicitly drawn in contemporary philosophy.

### On Normativity and 'Reason'

Contemporary philosophers sometimes call normative anything that comprises, as a matter of descriptive fact, a set of norms, rules, or precepts by which agents might be guided. Roman Catholicism, chess, English law, French etiquette, prudence, and morality are all "normative" in this sense insofar as they each prescribe various types of mental state or action. These codes are each, as a matter of descriptive fact, clearly *prescriptive*: they tell agents what to do or what attitudes to have. But it remains a further question whether one has any *reasons* or *ought* to fulfil the precepts of each code. This points to the second sense of 'normative': something is *genuinely normative*, one might say, insofar as its precepts provide genuine considerations that can help determine what one ought to do (or in what state one ought to be). If Roman Catholicism is genuinely normative, then not only does it prescribe that agents undertake some actions, they also have reasons to do so, and it might even be the case, if the reasons are decisive, that they ought to do so overall. Classical natural-law theorists such as Suárez made this point by distinguishing between precepts or so-called laws that merely provide a "rule" or "measure" and "prescribe a certain mode of action" but which are not binding, versus laws in the

<sup>50</sup> Hampton (1992).

“proper,” binding sense.<sup>51</sup> Following their cue, in this book I shall call a code or precept that is normative in the first, descriptive sense *prescriptive* and reserve the term ‘normative’ for the second, genuine sense defined in terms of reasons.<sup>52</sup>

‘Reason’ in the sense I have just used the word is (1) a count noun denoting considerations for and against doing something, or being in some state, with the power to help determine what one ought to do or what state one ought to be in overall. These genuinely *normative reasons* include epistemic reasons for belief, affective reasons for passions or conative attitudes, and practical reasons for action. But ‘reason’ can also be used as (2) a count noun to mean factors that help explain what individuals *actually* do or what state they actually are in. These are *explanatory reasons*. ‘Reason’ can also be used as (3) a count noun to mean an agent’s *motivational reasons*, which are factors that (a) help explain what individuals actually do or what state they are in, (b) in virtue of the fact they take those factors to be normative reasons for doing it or for being in that state. Motivational reasons thus compose a special class of explanatory reasons the concept of which presupposes the concept (but not the existence) of normative reasons.<sup>53</sup> I argue in [Chapter 1](#) that Hobbes used ‘reason’ in all three ways. He also used ‘reason’ as (4) a verb, to mean the *activity of reasoning* and, relatedly, as (5) a mass noun, to mean the *capacity for reasoning*, i.e., the “faculty” of reason. These are the two senses at stake in his well-known definition of “this word *Reason*” in [chapter 5](#) of *Leviathan*, according to which “REASON, in this sense, is nothing but *Reckoning* (that is, Adding and Subtracting) of the Consequences of generall names agreed upon.” This is a definition of the verb, not the count noun.<sup>54</sup> As we shall see in [Chapter 3](#), Hobbes also often used ‘reason’ (6) to mean agents’ *own* act of reasoning and/or subjective *judgement* or belief, resulting from the act of reasoning, about the normative reasons they have.

We should distinguish two further senses of ‘reason’: one Hobbes did not use, another he did. Consider the difference between a *source* of precepts and a mere collection or *articulation* of precepts. Roman Catholicism, chess, English law, French etiquette, etc., might each be a source of precepts: if chess is a source, then the precept requiring moving

<sup>51</sup> DLDL I.I.6.

<sup>52</sup> For these two senses of ‘normative’, see Parfit (2011: vol. 1, 144); Broome (2013: 11). Cf. Wright (1992: 15).

<sup>53</sup> On explanatory, motivational, and normative reasons, see Darwall (1983: 29–34); Smith (1994: 95–98); Schroeder (2007: 11–14); Crisp (2006).

<sup>54</sup> L 5.2: 64.

one's knight in an L-shape would not exist without the game.<sup>55</sup> Alternatively, each might be, rather than a source, a mere articulation of independently existing precepts, the way a textbook on international law may simply collect international legal precepts whose sources lie elsewhere. Regardless of whether some prescriptive code is a source or mere articulation of precepts, it remains a further question whether its precepts are genuinely normative. If a source is itself genuinely normative, then its precepts are normative too, and so *ground* reasons; if a mere collection articulates precepts that are themselves genuinely normative (whatever their source), then those precepts *articulate* reasons. By contrast, if a source is not normative, or a collection articulates non-normative precepts, then its precepts do not specify genuinely normative reasons: they are *merely* prescriptive. If Roman Catholicism were not a normative source of precepts for anyone, and the Roman Catholic precept forbidding contraception had no other normative source, then the precept would not be normative for anyone and so would fail to ground or articulate any normative reasons to avoid contraception. Alternatively, if the Roman Catholic code articulated only independently existing moral precepts, and if morality were a genuinely normative source, then although Roman Catholicism would comprise a mere collection rather than source of precepts, the precept against contraception would still be normative – but in virtue of being a moral, not religious, precept. Finally, if Roman Catholicism were an independent and normative source of precepts, then the religious precept would provide a normative reason to avoid contraception.

Philosophers have often used the term 'reason' as a mass noun to mean a distinct source of precepts in the sense just specified – the way that chess or English law might be a source. To distinguish this sense of 'reason' from others, contemporary philosophers have begun using the term 'rationality' in its stead. One might think that rationality *qua* source prescribes to agents certain kinds of consistency in their mental states. Perhaps rationality prescribes that one not hold contradictory beliefs or contradictory intentions; perhaps it prescribes that, if one believes that *p*, believes that *p* implies *q*, and cares about whether *q*, then one believe that *q*; and perhaps it prescribes that, if one believes that one ought to see to it that some event occur, and believes that whether it occurs depends on whether one adopts a certain intention, then one have that intention.<sup>56</sup> Or perhaps rationality

<sup>55</sup> See Broome (2013: 109–110, 116).

<sup>56</sup> For more precise formulations, see Broome (2013: 155–157, 170).

prescribes that one respond correctly to one's judgements or beliefs about the normative reasons one has;<sup>57</sup> that one do what one would have a sufficient reason to do if one's beliefs were true;<sup>58</sup> or that one form the beliefs and intentions one would have formed, given the available data, if one had engaged in processes of information gathering and reasoning that satisfy certain procedural criteria.<sup>59</sup> Yet whatever rationality *qua* source prescribes, it remains a further question (just as it remains a further question for Roman Catholicism or prudence) whether its precepts are *normative* – whether, for example, one has a normative *reason* to avoid contradictory beliefs or intentions.<sup>60</sup>

Clearly distinguishing the notion of a normative reason (expressed by a count noun) from the notion of reason or rationality as a source (expressed by a mass noun) enables one to ask whether one has a reason to fulfil the precepts of rationality. One might have such a reason either because rationality is itself a normative source, or because there are indirect (e.g. prudential) reasons for fulfilling its precepts.<sup>61</sup> It is also by distinguishing normative reasons from rationality as a source that we can distinguish two senses in which some action (or belief, intention, or agent) might be said to be “rational” or “irrational.” Sometimes what philosophers mean when they call an action irrational is that one has decisive reasons not to undertake it; any agent undertaking such an action would, to that extent, be irrational. But in a second sense, we might call an action irrational to the extent that undertaking it violates a precept of rationality. In this second usage, the fact that an agent believes, intends, or does things she has decisive reasons not to is insufficient to establish that she is irrational: she may simply be mistaken about what reasons she has. She would be irrational only to the extent she believes, intends, or does things in ways that violate the precepts of rationality *qua* source: for example, by having contradictory beliefs, intending to do things she judges she ought not to intend, etc.<sup>62</sup>

It may seem that Hobbes used the term ‘reason’ to mean a source of precepts when he wrote of “what is conformable, or disagreeable to Reason,” or what is “consonant to” and “against Reason,”<sup>63</sup> or of “precepts” or “dictates of Reason.” But I shall argue in [Chapter 1](#) that in such expressions Hobbes was using ‘reason’ instead as (7) a mass noun meaning the collection or *articulation* of the ensemble of *normative* precepts, i.e., of

<sup>57</sup> Scanlon (1998, 2014). Cf. Broome (2013: 88). <sup>58</sup> Parfit (2011: 34–37, 111–121).

<sup>59</sup> Sobel (2001). <sup>60</sup> Kolodny (2005); Broome (2007a, 2013). <sup>61</sup> Broome (2013: 195–196).

<sup>62</sup> Scanlon (1998: 25–32); Sobel (2001: 483). <sup>63</sup> L 15.40: 242; 30.6: 524; see also 15.10: 226.

the normative *reasons* agents have overall. Thus, in his debate with the Foole who defends injustice, when Hobbes wrote that attempting to take possession of sovereign power via rebellion “is against reason,”<sup>64</sup> he was claiming that one ought not to do it, that doing so runs against the balance of normative reasons. (I shall say that one “ought” to  $\phi$  when one has a decisive reason to  $\phi$  or the balance of reasons favours  $\phi$ -ing.) In saying that seeking peace – the first law of nature – is a precept of reason, Hobbes meant that rational agents have a normative reason to desire or be disposed to seek peace.

### The Overarching Argument

Chapter 1 begins by examining the metaethical interpretation of Hobbes as a *projectivist* and *error theorist* of normativity. On this reading, positive normative judgements and propositions attribute to objects irreducibly normative properties they do not have: positive normative judgements and propositions are invariably erroneous. This interpretation is motivated by the analogy Hobbes drew between judgements and propositions concerning normative properties and those concerning sensory qualities such as colour. On Hobbes’s view, our perception that objects possess a certain colour is erroneous: external objects are bodies whose particular shape and motions cause our sensory apparatuses to *perceive* them as having a particular colour, but colour in reality is a property of our perception, not of objects. Just as we erroneously project colour, so too do we project normative properties where there are none.

Two problems plague this interpretation. First, it implies that Hobbes took the positive normative propositions advanced in his own ethics – such as his claim that peace is good, or that reason prescribes the means of self-preservation – to be false. Second, the interpretation is not in fact supported by the analogy between sensory qualities and normative properties because, although Hobbes advanced a *projectivist* theory of sensory qualities, he did not advance an *error theory*. Hobbes’s natural philosophy demonstrates that the proposition ‘The object is green’ is always false if by ‘green’ is meant a sensory quality inhering in the object. But Hobbes’s point in debunking the pre-scientific account of sensory perception was reductively to analyze colour properties in terms of the extension and motion that do inhere in objects, and hence to provide a scientific basis for *reforming* the customary, pre-scientific meaning of terms such as

<sup>64</sup> L 15.7: 224.

‘green’. With the proper reforms in place, ‘The object is green’ can amount to the true claim that the object has natural properties disposing regular observers to *perceive* it as phenomenologically green.

The analogy might, however, support two other potential interpretations. One is an *ethical-naturalist reductionist* interpretation according to which normative properties and facts can all be reductively analyzed in terms of non-normative natural ones. With the appropriate reforming definitions in place, some positive normative propositions would be true. On a simple version of this interpretation, the proposition that peace and its relevant means are good amounts to the naturalistic claim that peace and its relevant means are desired, and the proposition that we have a reason to take the means to self-preservation amounts to the naturalistic claim that we are motivated to do so. This simple interpretation is not viable, however, because Hobbes clearly recognized that not everyone actually desires or is actually motivated to take the relevant means to peace. A more promising reductionist interpretation is that normative propositions describe what agents *would* desire or do upon reasoning properly.

The problem with the reasoning-based interpretation is that Hobbes did not use normative language in a solely descriptive way – to predict what agents would believe, desire, or do upon reasoning rightly. Normative language for Hobbes is inherently prescriptive as well: it is used to *tell* oneself or others what to believe, desire, or do. A noncognitivist *prescriptivist* interpretation, also suggested by Hobbes’s analogy between colour and normative terms, takes this aspect of normative language into account. But Hobbes could not have thought that the inherent function of normative language is solely prescriptive: he took normative claims to be truth apt. [Chapter 1](#) concludes that amongst interpretations of Hobbes’s ethics that deny irreducibly normative properties and facts, the most promising is a hybrid interpretation. On the hybrid reading, to say that agents have a decisive reason to seek to preserve themselves or to seek peace is (a) to predict that they *would* do so upon reasoning rightly, but also (b) to express one’s desire that they do so in *telling* them to do so.

In [Chapter 2](#) I argue that the hybrid metaethical interpretation of Hobbes is undermined by his account of practical *reasoning* and *counsel*: the interpretation fails properly to integrate the reasoning-based and prescriptivist elements composing it. Against the dominant interpretation of Hobbesian practical deliberation and action, I argue that for Hobbes human deliberation and action are reflectively responsive to both practical *reasoning* and judgements about *reasons*. When humans engage in

practical reasoning or give counsel to others, they do not merely (a) assent to or assert the proposition that some fact furnishes a sufficient reason to do something combined with (b) a decision or prescription to do it. Rather, the two elements are intrinsically linked: one decides to do or prescribes doing something *on the basis* of the proposition that there is a fact furnishing a sufficient reason to do it. The prescription to seek peace, for example, is supposed to be grounded in the fact that peace secures self-preservation and that self-preservation is good. Linguistic reasoning for Hobbes consists in reflectively *reasoning from* premises, i.e., inferentially drawing conclusions on the basis of prior propositions reasoners suppose or take to provide a reason from which the conclusion is drawn; similarly, good counsellors ground their advice in the reasons they cite for their advice. The hybrid interpretation severs this intrinsic connection between deciding or prescribing and reasons. [Part I](#) concludes that only an irreducibly normative interpretation of Hobbes's metaethics properly accounts for this aspect of Hobbesian linguistic practical reasoning and counsel.

[Part II](#) begins with the argument, in [Chapter 3](#), that Hobbes was neither a conative nor cognitive subjectivist about normative reasons. According to conative subjectivists, agents' normative reasons are all grounded in their desires; according to cognitive subjectivists, normative reasons all derive from (or are relativized to) actual beliefs about what reasons they have. But Hobbes was committed to the irreducible normativity of at least two objective sets of precepts: (a) precepts of instrumental transmission that prescribe desiring and taking the relevant means to one's normative ends; and (b) precepts of prudence that prescribe caring for one's ongoing good or felicity. The normativity of the prudential precepts implies that rational agents have *reasons of the good*. Hobbes's prudentialism, I argue, provides the normative foundation for natural laws.

Affective and practical normative reasons, moreover, are for Hobbes relativized not to what agents happen to think, nor to what actually turns out to be the case, but to the evidence epistemically accessible to them, i.e., to what they have a sufficient epistemic reason to believe. This is because the point of reasons-talk is to specify the mental states or actions for which rational agents are responsible, and no one can be normatively blamed for having failed to take into account epistemically inaccessible considerations. Thus when the laws of nature prescribe the means to self-preservation, they prescribe to agents the means they each can reasonably know or foresee will favour their self-preservation – where self-preservation denotes not bare survival, but the preservation of a life worth living.

Part II of the book ends, in Chapter 4, with an account of Hobbes's theory of the good. Hobbes was a hedonist about the good, and the ultimate good – felicity – consists in a life of ongoing pleasure and freedom from pain. Yet Hobbes departed from classical, Epicurean hedonism in two important respects. For Hobbes, neither felicity nor pleasure is a “final” end: since agents may derive pleasure from aiming at things other than felicity or pleasure, neither is the *aim* of all valuable action; and felicity does not mark the *termination* of desires. Hobbes took felicity primarily to consist, not in greater pleasures of *satisfaction*, or even in the objective fulfilment of desires, but in ongoing net mental pleasures of *anticipation* (engendered by the experience of repeated success in satisfying one's desires).

Chapter 4 closes with two complications concerning *instrumental* goods. The first concerns whether, on the reforming definition of ‘good’ Hobbes was implicitly proposing, particular actions or states are to be called instrumentally good for agents in virtue of the fact that they are means that will *actually* enhance felicity, or in virtue of the fact that they are means agents can reasonably expect will do so (even when, against the odds, they will not actually do so). I argue that on Hobbes's substantive theory, something is instrumentally good for agents in virtue of the fact that it will actually enhance their felicity, but that Hobbes's reforming, scientific definition reserves the term ‘good’ for those actions or states they can reasonably expect will do so. There is therefore a gap between Hobbes's substantive theory of the good and his account of the scientific meaning of ‘good’.

The second complication follows from Hobbes's recognition that whether something can be reasonably expected to enhance felicity may partly depend on whether it is called ‘good’. Hobbes recognized that language does not simply describe language-independent normative truths: it may interfere with the world it purports to describe. In *prescriptively subversive circumstances*, some things would foreseeably promote felicity even though the convention of calling them “good” would diminish it. Under such circumstances one has an epistemic reason to believe that the action or state is good, but a prudential reason not to call it “good.” In *deeply subversive circumstances*, the very fact of calling something “good” makes it such that the thing *itself* will foreseeably damage one's felicity. Hobbes's great political insight was that individuals in the state of nature face deeply subversive circumstances with respect to all the social means of self-preservation save one – namely, covenanting to enter a commonwealth – and that the subjects of a commonwealth, by

contrast, face prescriptively *self-fulfilling circumstances* in which agreeing to call some things “good” *makes* them the relevant means for securing peace. This is why what the sovereign declares to be the social means to peace and hence “good” will often turn out to *be* good.

Part III of the book turns to the second dimension of normativity in Hobbes’s ethics. In Chapter 5, I argue that Hobbes’s ethics is underwritten by a fundamental distinction between two types of normative reason corresponding to two distinct notions of normative responsibility and blame. The first dimension involves *reasons of the good* for which one is responsible in the sense that the mental states and actions to which they apply are *attributable* to one such that, should one fail to heed them, one is the proper object of criticism. One is the appropriate target of censure in this *critical* form in virtue of the fact that one is capable of reflectively grasping, responding to, and being guided by reasons. Being capable of answering to reasons does not imply, however, that one is obligated to others to respond or even listen to their criticism or advice. The second dimension of normativity, by contrast, involves *reasons of the right* for which one is responsible in the further sense of being *accountable* to others for them. Should one fail to heed reasons of this kind, those to whom the reasons are owed may justifiably criticize but also censure one in the stronger, second-personal and *reactive* sense, which intrinsically demands a normative response accounting for one’s actions and the attitudes they express – a response such as justification, excuse, or acknowledgement of wrong and hence apology, compensation, or redress. Holding criminals legally accountable for Hobbes takes precisely this reactive, communicative form: the function of punishment is to deter future violations, but it performs this prospective function by reiterating, and demanding that criminals acknowledge, the normativity of the violated law. Hence being accountable to others does not for Hobbes reduce to the non-normative, naturalistic relation of *being held accountable*: to be accountable is for others to have the normative standing to hold one accountable.

The laws of nature intrinsically articulate prudential reasons of the good, whereas the obligations acquired via renunciation, donation, and contract consist in reasons of the right. The latter constitute juridical obligations of *justice* owed to others. Against this it might be objected that Hobbes included justice or fulfilling covenants as the third law of nature, either because the obligations of justice derive from and are reducible to the purely prudential reasons agents have to fulfil covenants, or because natural laws are not merely prudential but are moral laws for which one is accountable to others. In Chapter 6, I argue that the question of whether Hobbes was

committed to a genuinely moral notion of obligation is obscured by the fact that Hobbes himself meant by ‘moral’ something different than what is typically meant by contemporary moral philosophers. By ‘moral reasons’ contemporary philosophers typically mean either reasons grounded in taking others’ interests impartially into account, or reasons owed to others. But by moral precepts Hobbes merely meant those that regulate *social interaction* affecting several individuals’ interests (especially precepts serving a shared good). In Hobbes’s sense, purely prudential precepts concerning the *social* means of self-preservation are “moral” precepts, even if they articulate self-regarding reasons wholly grounded in one’s own interests and are owed to no one. The natural laws prescribing peace and its means are moral precepts in Hobbes’s sense, but not in either of our contemporary senses.

Even though Hobbes did not use the term ‘moral’ in our contemporary senses, Hobbesian obligations are nevertheless moral in the contemporary accountability sense. This accountability dimension of normativity is neither derived from nor reducible to the first, attributability one. It is true that justice is *also* a law of nature, but that is simply because it is both the case that one has a directed obligation to fulfil covenants *and* that one has a prudential reason to be disposed to do so. Hobbes was therefore committed to the foundational normativity of not only a set of prudential precepts, but also of a set of precepts grounding juridical obligations. Yet there remains an intimate relation between the two dimensions of normativity. The imperatives of self-preservation (1) yield prudential reasons for *occupying* the second-personal standpoint from which obligations are acquired and for being disposed to *fulfil* those obligations, and (2) help to determine the *content* of obligations, both by grounding inalienable rights (to bare survival and to a life worth living) that negatively *limit* obligations and by positively supplementing their content. That prudential reasons shape obligations is explained by Hobbes’s *sign theory of consent*, according to which one actually consents to acquire obligations when others reasonably impute to one the will to do so, not when one *actually* wills to do so. In doing so, others must rely not only on semantic and syntactic conventions but also on the assumption that one acts in a minimally reasonable and prudent fashion. This is why obligations are partly shaped by prudential considerations. It is also why in a developed legal system the civil law inherently incorporates the laws of nature: judges must interpret the sovereign legislator’s will by imputing to him the prudential intention to conform to natural law.

In [Chapter 7](#), I argue that Hobbes's distinction between the two dimensions of normativity is premised on his distinction between rational agents and persons. Rational agents are those who, in virtue their capacity for reasoning, have normative reasons and hence are attribution-responsible for their intentional states and actions. But to be accountable-responsible to others, a rational agent must also be a person. A person – whether natural or artificial – is both a normative status and an artificial construct defined in terms of four roles: representer, representee, audience, and author or owner. To be a person, one must not only be duly *authorized* to be a person – which is a normative property – but also be *considered* a person – which is an artificial property dependent on rational agents' intentional states. The point of Hobbes's category of personhood is to show how the possibility of occupying the second-personal standpoint in which one becomes accountable to others depends on being recognized as a person by them.

The conclusion returns to the metaethical question of the nature of normativity. I conclude that for Hobbes normative properties are not *real* properties – they do not have causal standing – but are nevertheless the object of truth-apt, epistemically objective propositions. Because foundationally normative principles do not specify real, causally efficacious properties, their truth cannot be discovered via perception; rather, they are self-evidently known insofar as humans could not conceive themselves as reasoning agents and persons without also conceiving themselves to have normative reasons of the kind specified by those normative precepts. If Hobbes is a founder of naturalism in ethics, then he must be seen as an especially sophisticated kind: one who acknowledged normative truths and properties irreducible to non-normative natural ones.



PART I

*The Metaethics of Reasons*



## CHAPTER I

# *Naturalism*

Hobbes aspired to raze the teleological Aristotelian science of his day and to construct in its place a new scientific edifice on mechanistic, materialist metaphysical foundations.<sup>1</sup> He went so far as to claim that the only real entities are bodies the properties of which can all be reductively analyzed and redescribed in terms of extension and motion,<sup>2</sup> and was committed to a naturalist epistemology according to which all knowledge of the world originates in sensory perceptions caused by bodies in motion.<sup>3</sup> Commentators seeking to honour Hobbes's claim to be a systematic philosopher have therefore widely assumed that his mechanism rules out any distinct, irreducibly normative properties: they have invariably interpreted his ethics in ways that avoid positing normative properties above and beyond natural ones.

Hence on a widespread interpretation of Hobbes's ethics, there is nothing of irreducibly normative, reason-implying *value*, there are only objects of desire; and there are no irreducibly normative practical *reasons*, only mental states that are the natural causes of action. Hobbes encouraged such an interpretation by depicting the study of ethics and morals as the study of the physical basis of human psychology: in *Leviathan* he characterized "ETHIQUES" as a branch of "PHYSIQUES," and in *De Corpore* (1655) claimed that in the study of *moralia* we "consider the motions of the mind, namely, *appetite, aversion, love, benevolence, hope, fear, anger, emulation, envy, &c.*; what causes they have, and of what they be causes."<sup>4</sup>

There are at least three ways to flesh out this broadly naturalist interpretive approach to Hobbes's ethics. First, one might hold that if there are veridical normative judgements and true normative propositions, then they must be reductively analyzable as judgements and propositions concerning non-normative naturalistic properties and facts. If Hobbes took

<sup>1</sup> On Hobbes's mechanism and corpuscularianism, see Brandt (1927); Mintz (1969); Friedle (2012).

<sup>2</sup> LL 9: 125.    <sup>3</sup> L I.2: 22.    <sup>4</sup> L 9: 130–131; DC 6.6. Cf. AW I.1.

normative judgements to be reducible in this way, and thought that normative vocabulary could be used to make normative claims reducible to naturalistic ones, then he would be committed to some version of ethical-naturalist reductionism.

It is undeniable that Hobbes articulated his sciences of ethics in seemingly normative terms. He characterized self-preservation as individuals' greatest good,<sup>5</sup> and consequently "Peace" and "the way, or means of Peace" as a universal "Good" for all.<sup>6</sup> On the reductionist reading, however, Hobbes's point was not that self-preservation and peace are good in the sense that humans have irreducibly normative reasons to desire them, but that as a matter of empirical fact humans strongly desire their own preservation and, consequently, peace – or at least would do so in moments of clear-minded, informed reflection. Similarly, Hobbes defined the law of nature as "a Precept, or generall Rule, found out by Reason, by which a man is forbidden" to do what omits or is contrary to the requirements of his self-preservation. Natural laws comprise "dictates of Reason" and each "prescribeth" and "requireth" terms "by which we are obliged."<sup>7</sup> Yet on the reductionist reading, natural laws merely state natural causal facts about the means to self-preservation:<sup>8</sup> the point is that when rational agents who desire self-preservation know and adequately hold in view such laws, they will, thanks to deliberation or reasoning, be *caused* to take the means to their desired ends. Natural laws are descriptive or *predictive*: they predict that, under appropriate circumstances, rational agents *will* take those means – where the appropriate circumstances can be spelled out in purely descriptive, non-normative terms.

Second, one might hold that normative judgements consist in beliefs about the existence of irreducibly normative properties or facts, and that normative vocabulary is used to assert propositions corresponding to such beliefs, but that such beliefs and claims are systematically erroneous. If this were Hobbes's view, then he would be an *error theorist* of normativity.<sup>9</sup> His ethics would consist in giving an empirical account of how, given the nature of mind and language, human beings are erroneously led to project irreducibly normative properties onto objects and to assent to and reason from irreducibly normative (but false) propositions. On the error-theoretical interpretation, although Hobbes's own normative propositions concerning what is good or

<sup>5</sup> EL 17.14; DH 11.6.    <sup>6</sup> L 15.40: 242.    <sup>7</sup> L 14.3: 198; 15.41: 242; 15.26: 236; 15.1: 220.

<sup>8</sup> Hampton (1986: 47; 1992: 335–337); Cromartie (2011: 39–40); Schneewind (1998).

<sup>9</sup> See Mackie (1977).

prescribed by reason are strictly speaking false, his intention in advancing them would have been to produce the outcome he most desired – namely, peaceful coexistence – by persuading his fellow humans of his doctrine’s truth. He would do so by taking advantage of the fact, described in the psychological component of his ethics, that humans cannot help but to form normative commitments that shape their actions.

Finally, one might hold that normative judgements are not cognitive at all – that they do not represent things as having properties – and deny that normative language is used to make any claims concerning features of the world. Instead, one might hold that normative judgements are conative states – desires or resolutions to adopt some mental state or undertake some action or, more generally, to favour some state or event – and that normative discourse is used to prescribe favouring such states or events. If Hobbes thought that this is what normative judgements and discourse amount to, then his own normative language concerning what is good or dictated by reason would not commit his ethical doctrine to any properties and facts beyond non-normative, naturalistic ones. Such language would be used by Hobbes to prescribe favouring some states or events. Hobbes would be a noncognitivist *prescriptivist*.

The result of any of these three variants of the broadly naturalist approach is that for Hobbes seemingly normative judgements and propositions would be veridical only insofar as they reduce to natural ones; if they are meant to represent distinct normative properties or to express distinct normative claims, then they are erroneous, since there are no distinct, irreducibly normative properties or facts; and if normative judgements do not represent anything and normative utterances solely prescribe, then they are not truth apt at all and imply no positive commitment to normative properties and facts. The upshot would be that Hobbes subscribed to an essentially nihilist theory of irreducibly normative reasons. I take up each of these approaches in turn, beginning with the more radical, error-theoretical interpretation.

### 1.1 An Error Theory

On the error-theoretical interpretation, Hobbes’s science of ethics essentially consists in a descriptive account of the mind and its relation to the world – an account explaining why humans project normative properties onto objects, why such projections are erroneous, and why all positive normative propositions are false. This interpretation is primarily motivated

by Hobbes's analogy between sensory qualities, such as colour and odour, and the evaluative property of goodness.<sup>10</sup>

Hobbes developed his theory of colour perception partly in reaction to a late-Medieval scholastic theory according to which humans accurately apprehend external objects' properties because objects produce, via the sensory organs, a likeness of themselves in the mind. They produce this likeness by generating phantasms or "sensible species" in which their properties are reproduced or copied. Some believed that green objects propagate green-coloured images through the air and into the eye, producing in the mind phantasms that, like the objects themselves, are green; others held that the objects produce sensible species that resemble or pictorially represent the objects' properties. A distinct, higher intellectual faculty was in turn supposed to abstract from the objects' contingent, material properties to conceive "intelligible species" sharing in the objects' form. Thus, although the scholastics recognized a mediating entity between perception and physical objects, namely the phantasm or image, their epistemology had nevertheless taken the form of direct realism: we can directly perceive and grasp the (properties of) physical objects.<sup>11</sup> This was the theory of colour perception against which Hobbes advanced his own account; similarly, he developed his account of the mental representation of value partly in reaction to a scholastic (and Cartesian) theory according to which humans represent to themselves the objective goodness of things thanks to a higher and purely rational faculty of understanding or judgement, distinct from the imagination.

Hobbes heaped scorn on both of these traditional theories as so much scholastic claptrap.<sup>12</sup> Already in 1640, in his manuscript *Elements of Law*, he had reduced the faculties of mind to "two sorts" of mental "Powers," namely, the "Cognitive or Imaginative or Conceptive," on the one hand, and the "motive" or conative power, on the other. The former consists in the mind's ability to have "Images or Conceptions," which comprise

<sup>10</sup> The interpretation is suggested by Tuck (1989: 53–55; 1996: 180–181, 191) and most fully developed by Darwall (1998: 92–93; 2000: 330). Although in principle he leaves it open whether Hobbes's projectivism entails an error theory or is compatible with noncognitivism, Darwall (2000: 338) in practice attributes the error theory: Hobbes provided "an account of our normative thought and judgment" as consisting in judgements about how "we ought to act," namely, "as they [natural laws] dictate," but did not provide "an account of the normativity of the laws of nature"; there "can be no [such] account, because . . . there really is no such thing" as normativity.

<sup>11</sup> Tweedale (1990) and Lagerlund (2008). Aristotle did not use the expressions 'sensible/intelligible species', and it is controversial whether he held that objects' properties are literally reproduced in perception. The attribution is made by Sorabji (1974), rejected by Burnyeat (1992). Cf. Caston (2005).

<sup>12</sup> L 1.5: 24. Tuck (1988b); Pink (2004).

sensory perception and imagination, while the latter consists in the power to have “affections and Passions.”<sup>13</sup> Hobbes thus treated all conceptions or purely cognitive mental states as products of a single faculty – namely, sense or the imagination – thereby rejecting the scholastic and Cartesian distinction between a lower, passive, sensory faculty, which gives rise to images, and a higher, active, rational or intellectual faculty, which gives rise to ideas.<sup>14</sup> This in part reflects Hobbes’s nominalism: he took all conceptions to be of some particular object, and thus rejected the possibility of universal or abstract conceptions.<sup>15</sup> The passions, in turn, are conative mental states in the sense that they in part comprise an urge or disposition to act: “appetite” or “desire” when they dispose us to favour some object, “aversion” when they dispose us to disfavour it.<sup>16</sup> These “Passions of Man” are “the beginning of all his voluntary motions” or, equivalently, “the Indeavour, or internall beginning of animall motion.”<sup>17</sup>

Conceptions and passions both arise thanks to imperceptible, internal motions in our body. Hobbes located the seat of these two types of mental state in distinct parts of the body, speculating in *Elements* that conceptions arise from internal motions in our brain, and passions from the continuation of motion “to the Hart.”<sup>18</sup> According to *Leviathan*, sensory perception (“Sense”) arises when an “Externall Body, or Object” presently exerts pressure on one’s sensory organ, which produces internal motions carried by nerves “inwards to the Brain, and Heart,” which provokes a “counter-pressure” in the heart “Outward,” which in turn produces an outward motion carried by the nerves to the brain, giving rise to a mental “Representation or Apparence” of the object. Imagination, in the narrow sense, arises when such a counter-pressure is produced by internal motions persisting within one’s body despite there being no external pressure on one’s sensory organs from the object represented – as in memory, when “after the object is removed, or the eye shut, wee still retain an image of the thing seen,” or fiction, “as when from the sight of a man at one time, and of a horse at another, we conceive in our mind a Centaure.”<sup>19</sup> Passion, in turn, arises when this self-same counter-pressure of the heart produces an endeavour or motion either “toward something which causes it” (desire) or “fromward something” (aversion).<sup>20</sup>

Although sensory perceptions and imaginations are purely cognitive states, and passions are conative states, both types of mental state are

<sup>13</sup> EL 1.7–8; 6.9.    <sup>14</sup> O 5; DC 5.9.    <sup>15</sup> EL 5.6; see also L 4.6: 52; DC 5.8; 6.15.

<sup>16</sup> I follow Skorupski (2010: 83–84) in using the term ‘favouring’ as a disjunction of various pro-attitudes or acts, such as desiring, admiring, respecting, choosing, promoting, etc.

<sup>17</sup> EL 5.14; 7.2.    <sup>18</sup> EL 7.1.    <sup>19</sup> L 1.4: 22; 1.1: 22; 2.2: 26; 2.4: 28.    <sup>20</sup> L 6.2: 78.

*intentional* and hence cognitive states in the sense that they comprise mental representations of objects. Hobbes was an intentionalist: he took all mental states to be intentional states insofar as they are mental representations *of* objects *as* having some properties. Conceptions and passions are both *directed* on the intentional objects they are about and also represent these objects in a particular or *aspectual* way:<sup>21</sup> mental states “are every one of them a *Representation* or *Appearance*, of some quality, or other Accident of . . . an *Object*.”<sup>22</sup>

We can therefore distinguish five components of Hobbes’s account of mental states: (1) the *mental state* itself, for which Hobbes’s most general names were “*Representation* or *Appearance*” and “*seeming*, or *fancy*” (“*Apparitio*” and “*Phantasma*” in Latin); (2) the *intentional object* of the mental state, that is, what the mental state is about or represents to the mind; (3) the *content* of the mental state, comprising, for example, the aspect under which the object is represented (what the object is represented *as*); (4) the *material basis* of the mental state, consisting in internal bodily motions; and (5) the *causal mechanism* by which these internal motions arise.<sup>23</sup> In visual perception, for example, Hobbes named the mental state itself “Sense” and more specifically “Vision,” deemed the causal mechanism by which it arises to involve the pressure exerted on the eye by the represented object, and the content to consist “in a *Light*, or *Colour figured*,”<sup>24</sup> i.e., in a representation of the object as having the properties of luminous colour and extension. And he speculated that the material basis of sensory perception consists in a motion in the brain.

Hobbes therefore agreed with the scholastics that our sensory perceptions represent external objects as having certain properties or “accidents,” including sensory qualities such as colour or sound. But already in *Elements*, he asserted that the perception that these sensory qualities inhere in external objects is a delusion. We are disposed erroneously to *project* the properties of our own intentional states onto their objects: “As Colour is not inherent in the object but an effect thereof upon us caused by such motion in the object . . . So neither is Sound in the thing we hear, but in ourselves.”<sup>25</sup> More generally,

<sup>21</sup> Crane (2001). <sup>22</sup> L I.I: 22; cf. 3.II: 46.

<sup>23</sup> On a mental state, its content, and its material basis in Hobbes, see Darwall (2000). Darwall calls the material basis of mental states the “appearance substratum,” which is unfortunate because ‘substratum’ denotes a substance, whereas motion is, according to Hobbes, an “accident” of substantial entities, not substance. It is unclear whether Hobbes took mental states to be *caused by* or to *consist in* internal bodily motions (Abizadeh 2017b).

<sup>24</sup> L I.4: 22. See MDO. <sup>25</sup> EL 2.9. Tuck (1988a).

whatsoever Accidents or qualities our Senses make us thinke there be in the world, they are not there, but are Seeminges and Apparitions only. The things that really are in the world without us are those motions by which these Seeminges are caused. And this is the great deception of Sense.<sup>26</sup>

*Leviathan* reiterates the point: “All which qualities called *Sensible*, are in the object that causeth them, but so many several motions of the matter, by which it presseth our organs diversly.”<sup>27</sup> The perception of an external object as luminous or green is not caused by a luminosity or greenness subsisting in and emanating from the object, but by imperceptible motions in the object’s constituent parts perpetuated outwards through a medium such as air.<sup>28</sup>

When we turn to evaluative properties, we find Hobbes once again agreeing with the scholastics that the mind can represent objects *as good*. But again, he denied that our mental representations of objects as good are subject to the judgement of some higher faculty. Moreover, he drew an explicit analogy between our perception of sensory qualities and our conception of goodness in objects. The analogy arises in the context of a structural parallel Hobbes drew between purely cognitive mental states such as sensory perception, in which objects are represented as having sensory qualities, and conative mental states or passions, in which objects are represented as having value or disvalue.

Although Hobbes did not have a specific name for the internal bodily motions that constitute the material basis of sensory perceptions or imaginations, he did have a name for the motions giving rise to passions. When the internal bodily motions giving rise to a passion are towards its intentional object, the motions are called “desire” or “appetite”; when away from the object, “aversion.” Strictly speaking, therefore, ‘desire’ and ‘aversion’ are not names of the mental state itself, but of its material basis – although Hobbes in practice often used ‘desire’ and ‘aversion’ loosely to denote the mental state as well (and I follow him in this). As he made clear in *Leviathan*, the name of the mental state itself is delight or *pleasure* in the former case, and trouble, molestation, or *displeasure* in the latter: the motion, when towards an object, “is called Appetite,” but “for the appearance of it [is called] *Delight*, and *Pleasure*.”<sup>29</sup> And the intentional content of conative mental states – i.e., the aspect under which the intentional object of pleasure or displeasure is represented – is as having the evaluative

<sup>26</sup> EL 2.10. Brandt (1927); Zarka (1999: 35–38); Tuck (1988b). <sup>27</sup> L 1.4: 22–24. <sup>28</sup> MDO 1.2.

<sup>29</sup> L 6.9–10: 82.

property of being good or evil. It is in making this final point that Hobbes drew the parallel with the projection of sensory qualities:

As, in Sense, that which is really within us, is . . . onely Motion, caused by the action of externall objects, but in apparence; to the Sight, Light and Colour; to the Eare, Sound; to the Nostrill, Odour, &c: so, when the action of the same object is continued from the Eyes, Eares, and other organs to the Heart; the reall effect there is nothing but Motion, or Endeavour; which consisteth in Appetite, or Aversion, to, or from the object moving. But the apparence, or sense of that motion [i.e., the mental representation to which the motion gives rise], is that wee either call DELIGHT, or TROUBLE OF MIND . . . *Pleasure* therefore, (or *Delight*,) is the apparence, or sense of Good; and *Molestation* or *Displeasure*, the apparence, or sense of Evill. And consequently all Appetite, Desire, and Love, is accompanied with some Delight more or lesse; and all Hatred, and Aversion, with more or lesse Displeasure and Offence.<sup>30</sup>

It is this parallel between the projection of sensory qualities onto objects (in having sensory perceptions) and the projection of evaluative properties (in having passions) that furnishes the textual basis for interpreting Hobbes as an error theorist of normativity. The key elements of this interpretation are as follows. First, Hobbes advanced a projectivist theory of sensory qualities: in perceiving external objects we project sensory qualities onto them, but in reality they possess no such properties. (The only properties inhering in bodies are extension and motion.) Second, as a consequence, Hobbes advanced an error theory of sensory qualities: all of our conceptions of objects as being green are erroneous, and any proposition via which we assert that objects are green is false. Third, Hobbes advanced a projectivist theory of the good: in desiring objects we project the distinct property of goodness onto them, even though they do not possess such a property. Fourth, as a consequence, Hobbes advanced an error theory of value: all of our conceptions of objects as being good are erroneous, and any proposition via which we assert that objects are good is false. Fifth, this error theory of value is generalized to all alleged normative properties, conceptions, and propositions. In short, there are no objective normative properties, and any assertion that there is a normative reason to believe, desire, or do something is false. The science of ethics consists in describing

<sup>30</sup> L 6.9–11: 82. I read the ‘of’ in “of that motion” as possessive, and not an ‘of’ of intentionality. Hobbes did not here mean that one’s internal motion is the intentional object represented to the mind, but that the internal motion gives rise to a representation of some intentional object. It is of course possible mentally to represent one’s own internal motion, but this is not the representational content of all pleasure as such. (Holden’s (2016: 141, note 13) reading of this passage is mistaken.)

the psychological processes by which we are led erroneously to project objective normative properties onto objects and to assert their existence.

This reading confronts two major challenges. The first arises from the fact that Hobbes's science of ethics does not seem restricted to mere descriptions of human psychology. It also comprises apparently evaluative claims about the good – most prominently, that self-preservation is the individual's greatest good,<sup>31</sup> and that "Peace" and "the way, or means of Peace" are universally "Good" for all.<sup>32</sup> Moreover, Hobbes linked these evaluative propositions to a set of natural laws characterized as the core of "Moral Philosophy," and which consist in "dictates of Reason" prescribing to agents the relevant social means of self-preservation and therefore of peace. These "Laws of Nature, are good" even if not everyone desires to comply with them.<sup>33</sup> The error-theoretical reading implies that Hobbes ultimately viewed the central propositions of his own science of ethics as false: in reality, self-preservation and peace are not of any genuine normative value, and the laws of nature fail to articulate normative prescriptions linked to anything of value.

One should not easily concede any interpretation with such incredible implications, for it obliterates the gulf that Hobbes boasted separates his own ethical system from his hapless predecessors': previous moral philosophies are mired in absurdity and falsehood, while his own furnishes a true science on a par with geometry. Hobbes proclaimed in *De Cive* (1642, second edition 1647) that "If the Morall philosophers had as happily discharg'd their duty" as the geometers of old, and "the nature of humane Actions [were] as distinctly knowne, as the nature of *Quantity* in Geometricall Figures," then humanity would presently "enjoy such an Immortall Peace, that . . . there would hardly be left any pretence for war. But now on the contrary" there is nothing but disagreement amongst "the severall factions of Philosophers," which furnishes "so many signes . . . that what hath hitherto been written by Morall philosophers, hath not made any progress in the knowledge of the Truth." What previous moral philosophers lacked is a rigorous, scientific method: the source of their "unluckines should seem to be this; That . . . there is not one that hath used an idoneous Principle of Tractation." Hobbes himself, by contrast, had discerned the correct "Principle of Tractation" establishing ethics on solid and true foundations.<sup>34</sup> Whereas the "hermaphrodite opinions of morall Philosophers, partly right and comely, partly brutall and wilde," have been

<sup>31</sup> DH II.6.    <sup>32</sup> L 15.40: 242.    <sup>33</sup> L 15.40: 242; 14.3: 198; 15.41: 242; 15.40: 242.

<sup>34</sup> DCv Dedicatory.7–8.

“the causes of all contentions, and blood-sheds,” his own moral philosophy would furnish the only veritable basis for peace.<sup>35</sup> But the error-theoretical interpretation implies that, by his own lights, Hobbes’s own philosophy is no less ridden with falsehood and absurdity than the “Vain and false Philosophy” of Aristotle so acerbically mocked in *Leviathan*.<sup>36</sup>

One might concede all this, but argue that Hobbes articulated his sciences of ethics and politics as a purely strategic, rhetorical performance designed to secure the end he most desired: to live in peace.<sup>37</sup> After all, Hobbes explicitly avowed his “hope, that one time or other, this writing of mine, may fall into the hands of a Sovereign, who will consider it himselfe . . . and by the exercise of entire Sovereignty, in protecting the Publique teaching of it, convert this Truth of Speculation, into the Utility of Practice.”<sup>38</sup> Hobbes may have believed that, despite their falsity, if his normative propositions were widely believed, his most cherished desire would be realized.

This rejoinder has the merit of explaining why, despite Hobbes’s own pronouncements against the role of metaphors and other rhetorical and literary devices, he himself deployed them with frequency (and mastery).<sup>39</sup> But if this was Hobbes’s agenda, then we face an obvious question: Why did he (allegedly) publicly articulate an error theory when doing so risks undermining the persuasiveness of his substantive normative claims?<sup>40</sup> Note that the allegation does not merely obliterate the gulf Hobbes claimed separates his own ethics and politics from Aristotle’s “vain” philosophy. It also reduces Hobbes’s ethics to a mere description or expression of his own most cherished passion, which was precisely how Hobbes castigated the ancient Greeks for “Their Morall Philosophy,” which “is but a description of their own Passions.”<sup>41</sup> The implication of the error-theoretical interpretation is that Hobbes covertly took this to be true of his own moral philosophy as well.

The second major challenge is that neither Hobbes’s account of colour perception nor his analogous account of conceiving value supports an error theory. It is true that Hobbes advanced a *projectivist* theory of colour: we are disposed erroneously to project the representational content of our mental states onto their intentional objects. But he was not thereby committed to an

<sup>35</sup> DCv Preface.7.    <sup>36</sup> Quoting L 46. 33: 1090, but referring to chapter 46 in general.

<sup>37</sup> Tuck (1996: 197) suggests that for Hobbes rhetoric is “solely concerned with winning victory” in persuasion. For a reading of Hobbes’s works as rhetorical interventions, see Johnston (1986).

<sup>38</sup> L 31.41: 574.

<sup>39</sup> On Hobbes’s strictures against, and widespread use of, literary and rhetorical tropes and figures, see Kahn (1985); Skinner (1996).

<sup>40</sup> For an attempt to address this challenge, see Darwall (2000).    <sup>41</sup> L 46.11: 1058.

error theory of colour according to which all propositions such as ‘x is green’ are false. Hobbes was not claiming that sensory perceptions do not *represent* the properties of external objects; he was claiming they do so indirectly, in the sense that the properties they represent do not *resemble* the sensory content of mental states.<sup>42</sup> Moreover, we are able to know this fact. Hobbes had made this point in his very first published work, the *Objectiones* (1641) to Descartes’s *Meditationes*: “When I think of a man, I am aware of an idea or image made up of a certain shape and colour; and I can doubt whether this image is the likeness of a man or not.”<sup>43</sup> Thanks to science, we can know that properties inhering in external objects, and which cause us to perceive them as green, for example, consist in specific combinations of extension and motion. As the error-theoretical reading might acknowledge, one of the tasks of science, as Hobbes conceived it, is to demonstrate this fact.<sup>44</sup>

What the error-theoretical reading does not acknowledge is that another task of Hobbesian science is to reform language, at least as used within science, in light of the requirements and findings of scientific inquiry. Part of the scientist’s task is to propose *reforming definitions* so that we can use language more scientifically.<sup>45</sup> As Hobbes put it in *De Corpore*:

whatsoever the common use of words be, yet philosophers, who were to deliver their knowledge to others, had always the ability [*potestas*], and sometimes they both had and will have a necessity, of taking to themselves such names as they please for the signifying of their meaning [*sensus*].<sup>46</sup>

Hobbes had made the same point in *Leviathan*, arguing it is necessary for the scientist “to examine the Definitions of former Authors; and either to correct them, where they are negligently set down; or to make them himself.”<sup>47</sup> His own practice conformed to these strictures: Hobbes regularly provided scientific, reforming definitions for the terminological building blocks of philosophy. Scientific reasoning requires “apt imposing of Names.”<sup>48</sup>

<sup>42</sup> Tuck (1988a) makes the same substantial point, but (misleadingly) calls indirect representation “non-representation.”

<sup>43</sup> O 5. We can *visually perceive* something as having some property, while believing or knowing that it does not. When we are aware that the image we see is an illusion created “by glasses,” then “we know the thing we see, is in one place; the appearance, in another.” L 1.4: 24.

<sup>44</sup> Holden (2016: 131–132) is mistaken to think this distinguishes sensory properties from evaluative ones.

<sup>45</sup> On reforming definitions, see Brandt (1979); Railton (1989). <sup>46</sup> DC 2.4; cf. 6.15.

<sup>47</sup> L 4.13: 56. See L 34.1: 610.

<sup>48</sup> L 5.17: 72. For Hobbes’s project of redefining mathematical terms, see Jesseph (1999). Schaffer (1988: 286–287) mistakenly asserts that Hobbes denied to geometers “the right to define their own terms.” See Jesseph (1999: 140–141).

Once science teaches us the difference between the sensory properties we project onto external objects and the properties actually inhering in them, we can reform language for scientific purposes. With the proper reforms in place, 'x is green' would not mean that x possesses a property resembling the content of a normal observer's visual representation of it; rather, it would mean that normal observers *see* x as having a certain visual property, and that x possesses the dispositional property to cause normal observers to see it that way, i.e., it possesses a combination of extension and motion disposed to cause normal observers to perceive the object as "green" in the pre-scientific, phenomenological sense.<sup>49</sup> Once we distinguish between the pre-scientific and scientific meanings of 'green', the proposition 'x is green' can be given two types of interpretation: one corresponding to pre-scientific or unscientific (scholastic) accounts of colour perception, another to Hobbes's scientific theory. Whether the proposition 'x is green' is always false will thus depend on what we mean by it: if we use 'green' in the non-scientific sense, then it is always false; but if we use 'green' according to the reforming definition proposed by Hobbes's science, then it is sometimes true that x is green. This is why Hobbes could treat, as he did in the sixth chapter of *Elements*, the proposition "It is Greene" as potentially true: despite his projectivist theory of colour perception, and contrary to what the error-theoretical reading supposes, Hobbes did not think that the proposition 'It is Greene' is always false. Its truth depends on whether it is an instance of "the proper use of names in Language."<sup>50</sup>

The mistake behind imputing to Hobbes an error theory of colour judgements and propositions is the false assumption that projectivism rules out veridical cognitive judgements and true propositions. Hobbes's projectivism is a doctrine about the proclivity of sentient beings to perceive objects as having properties they do not have. His projectivism does not, however, entail that linguistic beings cannot reflect on and judge as erroneous their own projections, or that the propositions they use to articulate beliefs about external objects' properties are systematically false. To the contrary, his theory of colour is premised on the possibility of reflecting on and of judging erroneous such projections. Rather than an error theory, Hobbes was proposing a *reductionist* theory of colour: he

<sup>49</sup> Extension and motion are necessary accidents (LL 9: 125). In observers, an accident is "the manner [*modum*] by which any body is conceived," while in the object itself it is "that faculty of any body, by which it works in us a conception of itself." DC 8.2.

<sup>50</sup> EL 6.1. Hobbes was recounting the tale of the Duke of Gloucester who exposed as a liar a man claiming to have been blind since birth. See Shakespeare's *Henry VI*, part 2, act 2, scene 1.

sought reductively to analyze sensory qualities in terms of extended bodies in motion and to reform scientific language accordingly.

If Hobbes was not an error theorist of colour, then there is no textual basis on which to read him as an error theorist of normativity. The analogy he drew between sensory and evaluative qualities suggests instead two other potential views of normativity and normative discourse: reductionism or prescriptivism.

## 1.2 Descriptivist Reductionism

On the readings of Hobbes we are considering here, there are no distinct, irreducibly normative properties. We can conceive of agents as having irreducibly normative reasons, but all such conceptions are erroneous, and any corresponding propositions false. Yet on the reductionist reading, normative propositions can be used, thanks to Hobbes's reforming vocabulary, to articulate facts that are reducible to non-normative naturalistic ones. This reductionist reading makes three central claims about Hobbes's ethics: (a) all normative properties and facts are ultimately reducible to non-normative, naturalistic ones; (b) normative judgements are representational (they represent objects as having normative properties) and normative utterances are truth apt; and (c) with Hobbes's reforming definitions in place, normative propositions can be used to describe objects' normative properties and assert true facts. On this reading, Hobbes was a *reductionist* about normative properties and facts, a *cognitivist* about normative judgements and propositions, and a *descriptivist* about the meaning of normative language.

On a simple version of the reductionist interpretation, normative language is used to describe what agents actually believe, desire, or do. Such an interpretation is often defended especially on the strength of Hobbes's famous declaration, in [chapter 6](#) of *Leviathan*, that "whatsoever is the object of any mans Appetite or Desire; that is it, which he for his part calleth *Good*: And the object of his Hate, and Aversion, *Evill*."<sup>51</sup> Those who find evidence here for a reductionist Hobbes often take him to have been claiming here that 'good' just means *desired* – that to say "x is good" is to say "I desire x."<sup>52</sup> This definitional or "analytical" reading of Hobbes's famous declaration is supposed to support a reductionist interpretation

<sup>51</sup> L 6.7: 80.

<sup>52</sup> For the claim that Hobbes equated the meaning of 'good' with *desired*, see Watkins (1955: 140–141); McNeilly (1968: 119–212); Gauthier (1969: 7–8); Hampton (1992: 335); Tuck (1996: 181). The claim is widespread in the moral philosophy literature (Miller 2003: 13).

because the reason why he declared ‘good’ to mean desired is supposed to be that there can be no further (irreducibly normative) property for the term to denote: being good just is the natural property of being desired, such that true normative propositions turn out, upon analysis, to describe a state of affairs. Although those advancing this definitional reading sometimes say they are attributing a kind of “subjectivism” to Hobbes, it is important to see that the reading attributes to him an account of meaning or word use, and not any substantive theory of normative value.

At other times, commentators – sometimes the very same ones – who call Hobbes a subjectivist seem to attribute to him a substantive, *desire-fulfilment theory of value*, according to which what *makes* something normatively valuable for agents is the fact they desire it.<sup>53</sup> On this substantive theory, being desired and being good are distinct properties – the former is natural and non-normative while the latter is genuinely normative and reason-implying – and what makes something good for people is that they desire it. Whereas the former, merely “analytical” form of the desire theory of an agent’s good is compatible with the simple reductionist reading, the latter, substantive desire-fulfilment theory is not, because on the substantive theory, the fact that something is desired confers to it a further, genuinely normative property: precisely in virtue of being desired, it becomes *desirable*, and the agent acquires normative reasons to favour it. That an agent has a motivational reason makes it the case that he also has a normative reason, but this further normative fact is not reducible to the natural fact of having that motivational reason.

To evaluate the merits of these rival readings we must therefore keep apart two questions: What is the *meaning* of the word ‘good’? And: What *makes* something good (or: in virtue of what is it good)? Those who find in Hobbes’s famous declaration evidence of his reductionism often read it purely as a response to the first, semantic question, while those who read it in a substantively subjective way take him to have been answering the second, metaphysical question as well.

When his famous declaration is read in isolation, it is not immediately clear which questions Hobbes took himself to be answering. But once the passage is read in context, it becomes clear that these two questions could not have been the only ones. Whatever else he was doing here, he was at the very least answering a third, psychological question, namely: What *causes* people to think that something is good and/or to call it good? We know

<sup>53</sup> See Gauthier (1979); Hampton (1986: 27–42, 53); Ripstein (1987: 124); Meyers (2013: 273).

this thanks to the role the passage plays in setting up Hobbes's explanation of war.

War, the central character of Hobbes's tale of the human condition, is not explicitly introduced in *Leviathan* until the thirteenth chapter, after the reader has already been treated, in the first twelve chapters, to Hobbes's materialist account of human nature and artifice. The final three chapters of Part 1, "Of Man," end with an account of the considerations in favour of seeking peace (survival and contentment) and the available means for doing so (contract and authorization). But even prior to Part 1's pivotal thirteenth chapter, Hobbes had already foreshadowed the outbreak of war: after having encountered, in the five opening chapters, expositions on sensory perception, imagination, thinking, language, and reasoning, the reader finally stumbles upon the roots of war in precisely the passage from chapter 6 we are examining. Hobbes exposed the problem by first observing that, because people have different and mutating bodily constitutions, there could hardly be any single "Object" of which everyone would, at any given point in time, "consent, in the Desire."<sup>54</sup> It is immediately after this that he declared, in the sentence with which we are concerned, that people call the objects of their desire "Good."

Hobbes did not here spell out the dangerous social and political implications, but he would do so clearly later, at the end of chapter 15, where he summarized the whole problem to which his political philosophy would furnish the answer: (1) people have different and changing objects of desire and aversion and (2) are moved to call the objects of their desire "good" and aversion "evil"; therefore, (3) at any given point, they will disagree about what to call good and evil, "From whence arise Disputes, Controversies, and at last War."<sup>55</sup> Hence part of the point of Hobbes's famous declaration is to articulate the second premise of this central argument in his disagreement theory of war – a premise, not about the meaning of 'good', nor about the property in virtue of which things are good, but about what moves people to call things good.<sup>56</sup>

To argue that Hobbes was here describing what moves or *causes* people to call things good is not to imply that the passage tells us nothing about what 'good' means or what makes things good. But the analytical reading does not cohere with the central role Hobbes's declaration plays in his explanation of war: if calling things good is just to report on one's desires, then when one calls something good and another calls it evil there is no

<sup>54</sup> L 6.6: 80.    <sup>55</sup> L 15.40: 242.

<sup>56</sup> On Hobbes's disagreement theory of war, see Abizadeh (2011).

genuine disagreement and so no genuine *disputational* basis for conflict.<sup>57</sup> If when I say “Monarchy is good” I mean I desire it, and if when you say “Monarchy is evil” you mean you are averse to it, then in reporting our respective passions there is no dispute between us.

There could be a disputational basis for conflict if interlocutors used normative terms such as ‘good’ to make claims concerning putative properties not indexed to the speaker.<sup>58</sup> But this supposes that in practice the interlocutors use ‘good’ to name something other than their own passions, so that the definition of ‘good’ Hobbes was ostensibly providing would have to be a *reforming* definition departing from customary usage.<sup>59</sup> Now, there is no doubt Hobbes took ‘good’ to be a term for which moral philosophy must provide a scientific definition: “Morall Philosophy is nothing else but the Science of what is *Good*, and *Evill*.”<sup>60</sup> But he explicitly argued that terms that, beyond signifying “what we imagine” of the “nature” of the object we conceive, also signify the “nature, disposition, and interest of the speaker” himself, “can never be true grounds of any ratiocination” if, as a result of this dual function, they “are in the common discourses of men, of *inconstant* signification” and thereby cause protracted disputation and controversy.<sup>61</sup> In other words, if Hobbes was clarifying the meaning of ‘good’ in his famous declaration of [chapter 6](#), he was characterizing how ‘good’ is customarily used, not providing his reforming, scientific definition of the term. The characterization of ‘good’ in [chapter 6](#) is straightforwardly meant to be of a usage that *is* the basis of protracted disputation and controversy. His own scientific definition of ‘good’, by contrast, must serve as the basis for scientific consensus (even if the term continues to signify something about the speaker). The analytical reading according to which Hobbes was reducing the true meaning of ‘good’ to actually desired is therefore unviable and provides no support for Hobbes’s alleged reductionism about the good.

There is, moreover, decisive positive evidence against the reductionist interpretation. When Hobbes made normative claims, he clearly did not take himself to be describing or predicting what intentional agents will actually desire or do. He was perfectly aware that sometimes – indeed, often – we fail to be motivated to take the relevant means to our long-term good or most desired ends. There are, to be sure, *explanatory* reasons for why we fail to do so; for example, we may sometimes lack a decisive

<sup>57</sup> For this objection against descriptivism, see Moore (1922), and against such readings of Hobbes, Darwall (2000).

<sup>58</sup> Tuck (1996: 181). <sup>59</sup> Miller (2003: 185). <sup>60</sup> L 15.40: 242. <sup>61</sup> L 4.24: 62.

*motivational* reason to do so. But since Hobbes nevertheless maintained that it is a “dictate of reason” to do so, he could not have used that expression to describe what individuals actually desire or do. Even if some are disposed to violate the seventh law of nature against cruelty, it remains a precept of reason “by which we are obliged,” such that when we glory “in the hurt of another, tending to no end,” we do so “contrary to reason.”<sup>62</sup>

This discrepancy with what people actually desire, do, or are disposed to do is directly mirrored in what Hobbes said about the good – in particular in his distinction between an agent’s “Apparent or Seeming Good” and his *bonum verum*. An agent’s seeming good is whatever seems to him to produce the best consequences for him; and perhaps agents actually (are disposed to) desire or do whatever seems good to them. But if what is *truly* good for agents is distinct from their seeming good, then the former cannot consist in what they actually desire. Thus when Hobbes claimed that natural laws are “dictates of Reason” articulating the “Science of what is *Good*, and *Evill*,” and that “Reason . . . dictateth to every man his own good,” he could not have meant that reason or natural law dictates to agents their seeming good.<sup>63</sup>

Further evidence against the reductionist interpretation: Hobbes used ‘reason’ not just as a mass noun (in expressions such as ‘dictates of reason’ or ‘against reason’), but also as a *count* noun, in contexts in which he asserted agents have reasons *to* have certain attitudes or to undertake certain actions. To be sure, if Hobbes had used ‘reason’ as a count noun solely to mean explanatory or motivational reasons, no particular difficulty would arise for the reductionist reading. And in many instances when Hobbes used the term as a count noun, he simply meant an explanatory reason, as when, having asked how the Jews could “fall into” believing in the phenomenon of possession, he answered that “I can imagine no reason” other than “the want of curiosity to search naturall causes; and their placing Felicity, in the acquisition of the grosse pleasures of the Senses.”<sup>64</sup> He similarly deployed ‘reason’ to mean an explanatory reason when explaining the diversity of Jews’ responses to St. Paul’s efforts to prove, based on passages from “the Old Testament,” that “Jesus was the Christ”: “What was the reason, when they all beleevved the Scripture, that they did not all belevee alike; but that some approved, others disapproved the

<sup>62</sup> L 15.19: 232.

<sup>63</sup> L 15.4: 222; 15.40: 242; 6.7: 94; AW 30.25; DH 11.5; 12.1. Cf. EL 17.14; DCv 3.31–32.

<sup>64</sup> L 8.25: 118.

Interpretation of St. Paul that cited them . . . ?”<sup>65</sup> Nor was his use of the term in this sense restricted to explaining human attitudes or actions: “when a thing lies still, unlesse somewhat els stirre it, it will lye still for ever . . . when a thing is in motion, it will eternally be in motion, unless somewhat els stay it . . . the reason be the same, (namely, that nothing can change it selfe.)”<sup>66</sup> Hobbes of course also frequently used ‘cause’ to mean an explanatory reason, as when he wrote of the “cause of Sense,” the “causes of warre,” or the “cause of Absurd assertions.”<sup>67</sup>

Sometimes Hobbes also deployed ‘reason’ as a count noun in an explanatory sense to mean specifically motivational reasons, as when he wrote that “It belongeth therefore to the Office of a Legislator . . . to make the reason Perspicuous, why the Law was made.”<sup>68</sup> Describing how sovereigns reduced aristocratic offices to merely honorific titles, Hobbes wrote that “In processe of time these offices of Honour, by occasion of trouble, and for reasons of good and peaceable government, were turned into meer Titles” with “neither possession, nor command.”<sup>69</sup> Sovereigns have legislated and have weakened the aristocracy because, in each case, they took some considerations to be (normative) reasons to undertake those actions. Hobbes also invoked the notion of motivational reasons for belief: “some are moved to beleeve” that the scriptures are the word of God “for one, and others for other reasons.”<sup>70</sup> Unsurprisingly, he used ‘cause’ as a synonym for ‘reason’ in the motivational sense as well, as when he wrote of “the Causes, and Motives, for which it [a law] was made,” or asserted that “The finall Cause, End, or Designe of men” in joining a commonwealth “is the foresight of their own preservation.”<sup>71</sup>

But Hobbes also frequently deployed ‘reason’ as a count noun in a third way. This third usage is markedly distinct from the contexts in which he asserted that there exists a motivational or other explanatory reason for why agents believe, desire, or act as they do. Consider first the case of epistemic reasons to believe and affective reasons to have some passion. Writing of clerical claims to have transubstantiated a piece of bread into “a God, or a man, or both” – even though the object “looketh still as like bread as ever it did” – Hobbes insisted “there is no reason for any man to think it really

<sup>65</sup> L 42.32: 808.

<sup>66</sup> L 2.1: 26. For further examples of ‘reason’ *qua* explanatory, see L 6.53: 92; 12.11: 170.

<sup>67</sup> L 1.4: 22; 11.4: 152; 5.10: 70. <sup>68</sup> L 30.22: 542; see 11.4: 152. <sup>69</sup> L 10.52: 148.

<sup>70</sup> L 33.21: 604.

<sup>71</sup> L 30.20: 542; 17.1: 254. For another example of ‘final cause’ used in this way, see L 26.21: 430. For other uses of ‘cause’ to mean motivational reason, see L 26.10: 420; 42.56: 836; 42.131: 922; 43.12: 938; 44.1: 956. Hobbes used ‘just cause’ or ‘reasonable cause’ to mean a motivational reason backed by a normative reason in L 7.7: 102; 15.16: 230; 22.29: 370.

done; nor consequently to fear” the cleric.<sup>72</sup> The implication is that even if some *do* “think it really done” or *do* fear the cleric, they have “no reason” *to* do so – in some sense of ‘reason’ distinct from the reasons explaining their having done so. Elsewhere Hobbes warned “there is no more reason to expect good Advice from the rich, or noble, in matter of State, than in delineating the dimensions of a fortresse,” and wrote more generally of “the truth, or probability of his [a counsellor’s] reasons, and of the grounds of the advise he gives.”<sup>73</sup> He also disparaged beliefs unsupported by epistemic reasons, such as beliefs supposedly supported by supernatural inspiration: for a person to “say he speaks by supernaturall Inspiration, is to say he finds an ardent desire to speak, or some strong opinion of himself, for which hee can alledge no naturall and sufficient reason.”<sup>74</sup>

Next, consider practical reasons to act. Hobbes’s second law of nature prescribes to individuals that they lay down, if others also will, their “right to all things.” Hobbes qualified the precept by saying, “But if other men will not lay down their Right, as well as he; then there is no Reason for any one, to devest himselfe of his: For that were to expose himselfe to Prey, (which no man is bound to) rather than to dispose himselfe to Peace.”<sup>75</sup> Hobbes was not here predicting that no one will actually expose himself to prey. The same holds for his eighteenth law of nature, which prescribes unbiased arbitrators: Hobbes was perfectly aware that sometimes those who select arbitrators might lack sufficient motivational reasons to reject biased arbitrators; yet he claimed that under the right circumstances they nevertheless have a reason to do so: “For the same reason no man in any Cause ought to be received for Arbitrator, to whom greater profit, or honour, or pleasure apparently ariseth out of the victory of one party, than of the other.”<sup>76</sup>

Hobbes even deployed ‘cause’ as a synonym for this third, distinct sense of ‘reason’ as a count noun. While the use of ‘cause’ in an explanatory sense is familiar to readers today, its use in the justificatory sense is perhaps less so. But it is not entirely alien: our word ‘because’, which can be used in both explanatory and justificatory senses, reminds us of this older sense. When Hobbes wrote “there is no cause to doubt, but that the seed of *Religion*, is also onely in Man,” he did not mean that no one has been caused to doubt it; he meant that there is no epistemic reason *to* doubt that the genesis of religion lies in human nature.<sup>77</sup> Similarly, when he

<sup>72</sup> L 37.13: 694.    <sup>73</sup> L 30.25: 546; 25.15: 410.

<sup>74</sup> L 32.6: 580. For epistemic reasons, see DCv 14.13. For affective reasons, see L 30.11: 528.

<sup>75</sup> L 14.5: 200.    <sup>76</sup> L 15.32: 238; see 25.2–3: 398; 25.10: 404.

<sup>77</sup> L 12.1: 164. For another example, see R&C.11: 1137.

contemplated the deposition of King Chilperic III by Pope Zachary “for no cause,” he was contemplating an action the pope lacked a sufficient reason *to* undertake.<sup>78</sup> When he wrote that “Leagues of Subjects . . . are in a Common-wealth . . . for the most part unnecessary, and savour of unlawfull designe; and are for that cause Unlawfull,” he meant that commonwealths have a decisive reason *to* outlaw such leagues.<sup>79</sup> And when he wrote that subjects who actually follow the example of idol worshippers “had no cause to follow such example,” he meant that, while there evidently exists an explanatory reason for their action, subjects lack a sufficient reason *to* parrot idolaters.<sup>80</sup> Again, to write of “causelesse fears” is not to write of fears whose genesis has no causal explanation, but of fears that agents lack any sufficient reason to have – in some sense distinct from the motivational or other explanatory senses of the term.<sup>81</sup>

The use of ‘reason’ as a count noun in this distinct, third sense is mirrored by Hobbes’s use of the term ‘ought’. The way Hobbes linked the two terms strongly suggests he did not use either to describe what people actually or even tend to believe, desire, or do. Hobbes wrote, for example: “nor by the same reason ought they to believe, that the Government is of one kind, when they like it, and another, when they dislike it, or are oppressed by the Governours.”<sup>82</sup> This parallels his linkage of ‘reason’ and ‘ought’ in the affective and practical context of the eighteenth law of nature, where Hobbes claimed that “For the same reason no man” with a vested interest “in any Cause ought to be received for Arbitrator.”<sup>83</sup> Hobbes’s use of ‘ought’ is central to his articulation of the laws of nature, which all derive from the “precept, or generall rule of Reason, *That every man, ought to endeavour Peace, as farre as he has hope of obtaining it.*”<sup>84</sup> He explicitly distinguished utterances deploying ‘ought’ from descriptive and explanatory propositions, insisting that “though in all places of the world, men should lay the foundation of their houses on the sand, it could not thence be inferred, that so it ought to be.”<sup>85</sup> Although fear frequently explains people’s actions, “not every Fear justifies the Action it produceth.”<sup>86</sup> Describing or explaining what is actually the case differs from justifying what ought to be the case.

<sup>78</sup> L 12.32: 186.    <sup>79</sup> L 22.29: 370.    <sup>80</sup> L 45.27: 1038.

<sup>81</sup> L 8.20: 112. For further examples, see L 2.7:32; 14.16: 206; 14.19: 210; 36.20: 678; 44.37: 1006. Cf. Grotius on “justificatory causes,” JBP II.22.1.

<sup>82</sup> L 19.2: 284.    <sup>83</sup> L 15.32: 238; cf. 25.10: 404.    <sup>84</sup> L 14.4: 198–200.

<sup>85</sup> L 20.19: 320–322. This speaks against Darwall’s (1992: 163) claim that ought is equivalent to a decisive motive for Hobbes.

<sup>86</sup> L 27.20: 464.

### 1.3 Reasoning-Based Descriptivist Reductionism

Hobbes's naturalism therefore cannot take the form of reducing normative judgements and propositions to representations or assertions concerning what agents actually believe, desire, or do. The challenge for the reductionist reading is to analyze and explain Hobbes's (a) distinction between apparent and true good, (b) use of 'reason' as a mass noun in expressions such as 'dictates of reason', and (c) third use of 'reason' as a count noun and of 'ought' – without reducing these to a simple description or prediction of what agents will actually (or tend to) believe, desire, or do.

A potential response to this challenge stems from Hobbes's observation that we often desire what would bring short-run pleasures, but upon reasoning properly would most desire what conduces to ongoing pleasure in the long run. Consider the following passage from *Elements*:

Every man by naturall passion, calleth that good which pleaseth him for the present, or so farre forth as he can foresee . . . And therefore he that foreseeeth the whole way to his preservation, which is the end that every one by nature aimeth at) must also call it good, and the Contrary, Evill. And this is that good, and evill, which not every man in passion calleth soe, but all men by reason. And therefore the fulfilling of all these Lawes is good in reason . . . And from hence cometh that distinction of *Malum pœna*, and *malum Culpæ*, for *Malum pœna*, is any paine or molestation of minde whatsoever. But *Malum culpæ* is that action which is contrary to reason and the law of nature.<sup>87</sup>

On a more sophisticated reductionist interpretation, Hobbes was here claiming that, while not all rational agents will in fact desire the means to self-preservation or fulfil the laws of nature, all *would* do so if (and because) they engaged in proper *reasoning*: “not every man in passion” calls the means to his preservation good, but “all men by reason,” that is, those who *have properly reasoned*, do. Similarly, while *malum pœna* comprises actions that agents actually do not desire or undertake because they perceive them and their consequences to be unpleasant, *malum culpæ* comprises actions that agents would not desire or undertake upon reasoning properly – because they would come to perceive their consequences as, on balance, unpleasant in the long run.<sup>88</sup>

This reasoning-based reductionist reading suggests that Hobbes used normative-sounding vocabulary to articulate a prediction about what rational agents would desire or do under descriptively specifiable

<sup>87</sup> EL 17.14.    <sup>88</sup> Hobbes ruled out *akrasia*. LN §23; L 6.53: 92.

circumstances.<sup>89</sup> Thus if ‘apparent good’ means whatever one actually desires, such as what yields short-run pleasure, then ‘true good’ would mean what one would desire (or pursue) upon proper reasoning or deliberation, i.e., one’s long-run balance of pleasure.<sup>90</sup> As Hobbes put it in *De Homine* (1658), “although the true Good is to be sought in the foreseeable long term, which is the work of Reasoning, Appetite seizes a present Good,” and the passions often “impede right reasoning by militating against the true Good in favour of an Apparent and immediately Present good, which frequently ends up (weighing carefully everything adjoined to it) being Evil.”<sup>91</sup> So when Hobbes wrote in *De Cive* that “They therefore who could not agree concerning a present, doe agree concerning a future Good [namely peace], which indeed is a work of Reason; for *things present* are obvious to the sense, *things future* to our Reason only,”<sup>92</sup> he was not claiming humans have an irreducibly normative reason to agree that peace is good. He was claiming that upon reasoning properly they *would* agree that peace is good, because it is via reasoning that humans come to see long-term consequences.

The question is whether the concepts of a “dictate of reason,” of what is “against” or “contrary to reason,” and of what one has a reason or ought to do could similarly be reductively analyzed. Take Hobbes’s use of ‘reason’ as a mass noun in expressions such as “dictate of reason.” One way to interpret ‘reason’ here is in irreducibly normative terms, taking it to mean the *collection* or articulation of the ensemble of normative precepts, i.e., of the normative *reasons* agents have overall. A dictate of reason would be one of the normative precepts composing this collection – a normative reason agents have. And to write of “what is conformable, or disagreeable to Reason” or what is “consonant to” and “against Reason” would be to write of what conforms to or goes against the balance of these normative reasons overall.<sup>93</sup>

The reductionist reading must provide an alternative analysis. One way to drain the normative significance out of these notions and expressions would be to take up John Broome’s suggestion. Broome suggests that when early-modern philosophers wrote that something is “against reason,” they were using ‘reason’ as a mass noun to mean not (the balance of) normative reasons, but a *source* of precepts – in the same way contemporary

<sup>89</sup> Cromartie (2011). This reductionist reading is also suggested by Hampton (1986: 38–40); Darwall (1995). See Peters (1956: 164); Forsberg (1990).

<sup>90</sup> Hampton (1992: 336). <sup>91</sup> DH 12.1. <sup>92</sup> DCv 3.31.

<sup>93</sup> L 15.40: 242; 30.6: 524. Cf. “the Conformity, or Inconformity to Reason . . . of particular Actions.” L 15.10: 226.

philosophers use ‘rationality’. It is then a further question whether this source and its precepts are genuinely normative. Broome’s example is David Hume’s famous statement that “‘Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger.”<sup>94</sup> Broome argues that Hume was not saying that there are no normative *reasons* against having such a preference, but merely that having the preference would not violate any precept of rationality. He suggests that the use of ‘reason’ as a mass noun to mean (decisive) normative reasons – as when philosophers write that an agent “has reason to” or “has most reason to” – is a twentieth-century innovation.<sup>95</sup> If Broome is right, then when Hobbes wrote of “what is conformable, or disagreeable to Reason,” or what is “consonant to” and “against Reason,” and when he called the laws of nature “precepts” or “dictates of Reason,” he would have been using ‘reason’ to mean rationality – or, as Hobbes himself would have put it, “the Method of Reasoning” or rules of logic<sup>96</sup> – as an independent source of precepts. If so, in these contexts the mass noun ‘reason’ would remain compatible with a reductionist analysis: the term would not itself be reason-implicating in an irreducibly normative sense.

Broome’s suggestion about the early-modern period is mistaken, however. Hobbes, at least, did not use ‘reason’ in “dictate of reason” to mean a source of normative precepts (consisting in, for example, rules of logic). Hobbes’s word choices furnish sufficient textual evidence against the Broomean interpretation. Consider Hobbes’s debate with the Foole who says “there is no such thing as Justice” and thus naysays the third law of nature prescribing fidelity to covenants. When Hobbes had the Foole say that in some circumstances violating covenants is “not against Reason,” he was simply rephrasing the latter’s contention that in such circumstances “there could be no reason, why every man might not do” such a thing.<sup>97</sup> Even if the former use of ‘reason’ is ambiguous, in the latter case he was clearly using the term as a count noun – to mean agents’ *reasons to* act.

Consider also Hobbes’s use of expressions such as ‘dictate of reason’. When Hobbes elaborated on what makes something a civil or prophetic law, he pointed to the precept’s source, such as the authoritative will of the sovereign or of God. But when he elaborated on what warrants calling a precept a precept *of reason*, he pointed not to its *source*, but to the *activity* of discovering via reasoning: a precept is “of reason” in virtue of the fact that it is discoverable only via the exercise of one’s faculty of *reasoning*.

<sup>94</sup> THN II.3.3.6.    <sup>95</sup> Broome (2013: 47, 194–195).    <sup>96</sup> L 46.11: 1060; DC 6. Q: 46/61.

<sup>97</sup> L 15.4: 222.

Thus to say the law of nature is a “precept, or generall rule of Reason” is to say it “is a Precept, or generall Rule, found out by Reason.”<sup>98</sup> There is no suggestion that natural laws are “dictates of Reason” in virtue of being precepts whose source is “rationality” (or, as Hobbes might have put it, the rules of logic): to say they are dictates of reason is just to say that “there is no other way to know” them than “by naturall Reason”<sup>99</sup> and that they are the “Conclusions”<sup>100</sup> or discoveries of reasoning rightly.

There remains, however, another interpretive possibility compatible with a reductionist analysis. The Broomean reading is not viable because for Hobbes something is correctly called a dictate of reason insofar as it is the *outcome* of right reasoning. But there are two potential interpretations of such an outcome. On the irreducibly normative interpretation, the outcome of right reasoning is the *discovery* of genuinely normative reasons. But a partisan of the reasoning-based reductionist interpretation might claim the outcome is not ultimately anything normative: it is just some particular attitude or action of the reasoner. To say that reason dictates some desire or action here is just to say that upon right reasoning one would acquire the desire or undertake the action in question. To dictate something here means to cause it. Normative propositions are used to describe non-normative, naturalistic features of the world.

The same analysis can even be applied to ‘reason’ used as a count noun in the third, distinctive way and to ‘ought’: for Hobbes to say that all rational creatures have a decisive reason or “*ought to endeavour Peace*” is to say they would do so upon reasoning rightly (even though they might not actually do so); it is not to say they have a decisive reason or ought to do so in an irreducibly normative sense. Similarly, for him to write “there could be no reason, why every man might not do” something is just for him to say that no agents *would* do it upon reasoning rightly. The descriptivist reasoning-based reading proposes reductively to analyze all of Hobbes’s apparently normative uses of ‘reason’ (whether as a mass or count noun) and ‘ought’ in terms of the more fundamental notion of right reasoning. It reduces all his normative-sounding talk to causal predictions.

The reasoning-based interpretation can also provide an account of how the use of such normative expressions gives rise to genuine disagreements. On this analysis, when I call some object good, in the sense that it is good *for me*, I assert that I would desire it upon right reasoning. But if I claim something is good in the sense that it is a common good *for you and me* – which is what I implicitly do when I call some set of actions virtuous –

<sup>98</sup> L 14.3–4: 198–200.    <sup>99</sup> L 31.33: 568.    <sup>100</sup> L 15.41: 242.

then I am asserting that both you and I would desire it upon right reasoning. By contrast, if you claim the same object is a common evil, then you are asserting that both you and I would be averse to it upon reasoning rightly. This yields a genuine disagreement: we are making conflicting predictions about each other. Similarly, if I claim that your undertaking some potential action is against reason, and you claim it is dictated by reason, then I am saying you would not, and you are saying you would, undertake the action upon reasoning rightly. This again is a genuine disagreement. Finally, such an account of what ‘good’ and ‘dictate of reason’ mean can serve as the basis for philosophical consensus precisely because and insofar as each rational agent could discover via right reasoning what each would in fact desire or do upon reasoning rightly.

It might be objected that insofar as this reading appeals to the notion of *right* reasoning, it covertly relies on, and implies that Hobbes’s normative-sounding claims presuppose, a genuinely normative property that cannot be reductively analyzed in purely non-normative, naturalistic terms. Such an objection can be readily met, however. As I show in [Chapter 2](#), for Hobbes, reasoning in its paradigmatic, linguistic form consists in the cognitive activity of drawing inferences from premises in accordance with and guided by rules or precepts of logic that constitute linguistic reasoning as a specific activity. One engages in linguistic reasoning insofar as one intends to follow what one takes to be these rules and so is guided by them. One reasons erroneously insofar as one misinterprets or misapplies the constitutive rules; one reasons rightly insofar as one correctly interprets and applies those rules. Hence right reasoning, in its linguistic form, certainly is “normative” in the descriptive sense that, as a matter of fact, it is constituted by rules. But a further question remains whether it is normative in the *genuine* sense that one has a normative reason to reason properly, and whether the prescriptive rules that constitute right reasoning are genuinely normative. (The question of whether one ought to engage in reasoning or be guided by logical rules was intelligible to Hobbes and his contemporaries: puritan enthusiasts claimed that, because we could not know God’s will via reasoning, we should rely on inspiration instead.) The reductionist can appeal to the descriptive sense of right reasoning and distinguish on that basis what agents actually desire or do from what they would desire or do upon reasoning rightly. It is in terms of this descriptive notion of right reasoning that the reasoning-based interpretation reductively analyzes the apparently normative uses of ‘reason’.

We should therefore be perfectly clear about what this reasoning-based interpretation presupposes: that there is nothing genuinely and irreducibly normative about reasoning and even right reasoning. There is no irreducible normative reason to engage in right reasoning, and the fact that one *would* believe, desire, or do something upon properly reasoning is not an irreducible normative reason to do so. Such alleged facts are reducible to purely non-normative and natural ones: they constitute a prediction about what one would do under some descriptively specified circumstances. When agents engage in reasoning or deliberation, they might *take* themselves to have irreducibly normative reasons to assent to some proposition, desire something, or undertake some action, and they may use normative vocabulary in a pre-scientific way to assert they have such reasons, but in reality they do not, and Hobbes was proposing to reform normative vocabulary to reflect this. If agents understand what they are doing when they reason, they will know that they are simply trying to cause the beliefs, desires, or actions that would result from reasoning rightly.

#### 1.4 Noncognitivist Prescriptivism

The reasoning-based reading shares three features with the simpler reductionist reading. First, it is *reductively naturalist*: it reduces all normative properties and facts to non-normative, naturalistic ones. Second, it is *cognitivist*: it takes normative judgements to be representational (insofar as they represent things as having certain properties) and normative utterances to be truth apt. Third, it is *descriptivist*: it reduces potentially true normative propositions to purely descriptive propositions.

The problem is that no descriptivist analysis of normative language can do justice to Hobbes. On any purely descriptivist interpretation, when Hobbes claimed that something is “good” or a “precept of reason,” or that one has a “reason” or “ought” to do something, he not only failed to say anything irreducibly *normative*, he also did not *commend* or *prescribe* anything to anyone. To say that others would do something upon reasoning rightly is merely to make a prediction; one does not thereby *commend* (praise, or laud) them for reasoning rightly or for actually doing what they would have done if they had reasoned rightly; nor does one thereby *tell them* to reason rightly or to do what they would have done if they had reasoned rightly. Yet Hobbes clearly thought an essential function of normative language is to commend and prescribe, and he used normative language with laudatory and prescriptive illocutionary force. Purely descriptivist readings cannot explain how, according to Hobbes, normative

terms are used in subjective discourse to *guide* deliberation and reasoning and in intersubjective discourse to *commend* and *advise*.

Consider purely descriptive utterances first. Normally when one asserts a proposition in a descriptive way, one *signifies* one's belief that the proposition is true. To signify something in Hobbes's technical sense is just to indicate or provide grounds for inferring its presence. Hobbes's examples: thick clouds signify rain for regular observers; names signify conceptions for language users.<sup>101</sup> Thus the word 'rain' uttered in a grammatical sentence signifies the presence in the speaker's mind of a conception of rain. When Hobbes wrote, "The source of every Crime, is some defect of the Understanding,"<sup>102</sup> he was (according to his own philosophy of language) signifying his belief to that effect. But while signifying one's beliefs is a feature of uttering propositions, it does not constitute such utterances' inherent function or illocutionary force: the function of Hobbes's utterance is to assert that the source of every crime is some defect of the understanding, not to describe his own mental states. Its function is to assert the belief's propositional content, not to report that he has the belief. Of course, one can utter propositions that, in addition to signifying one's beliefs, also serve to describe or report on them, such as when in *Anti-White* (1642/43) Hobbes reported that "I incline to the opinion that there can be no true proposition about the nature of God except this one: God is."<sup>103</sup> But asserting that one has such-and-such beliefs is a special case uncharacteristic of assertoric speech in general.

Hobbes clearly thought that normative speech acts signify speakers' *passions*. Whatever else they do, "*Good*, and *Evill*, are names that signifie our Appetites, and Aversions."<sup>104</sup> This distinguishes normative from non-normative speech, because the latter can exclusively signify purely cognitive mental states (such as opinions). The question is whether there is also anything distinct about the function or illocutionary force of normative speech acts or whether, as descriptivists maintain, normative utterances are used purely descriptively to assert the truth of a proposition and hence have the same underlying illocutionary force as non-normative utterances.

Consider evaluative speech. To value something, conceive it as good, or judge that it is good is what Hobbes called *honouring*: "Honour consisteth in the inward thought, and opinion of the Power, and Goodnesse of another," and "To Value a man at a high rate, is to *Honour* him." Since

<sup>101</sup> EL 4.9; L 3.8: 44; DC 2.2. On conventional signs, see DCv 15.16; L 2.10: 36; DC 2.2. Cf. Abizadeh (2015).

<sup>102</sup> L 27.4: 454.   <sup>103</sup> AW 35.16.   <sup>104</sup> L 15.40: 242.

the “manifestation of the Value we set on one another, is . . . commonly called Honouring” as well, the term may also be used loosely to name the mental states’ external signs. But strictly speaking, Hobbes called speech acts that outwardly express honour ‘praise’: “The forme of Speech whereby men signify their opinion of the Goodnesse of any thing, is PRAISE.”<sup>105</sup> Thus for Hobbes, although evaluative speech acts *signify* a cognitive mental state (namely, an opinion about an object’s goodness), their function cannot – as descriptivists assume – merely consist in describing objects in asserting that they possess some evaluative property. Their illocutionary force must consist in commending or praising.

Partisans of the reasoning-based reading may respond by applying to ‘praise’ the same reductive analysis they apply to ‘good’, ‘against reason’, and ‘ought’. They might insist that for Hobbes to “praise” is not actually to praise anything: it is solely to make a truth-apt, descriptive claim, namely, the prediction that others would favour the object of “praise” upon proper reasoning. The problem with this response is that Hobbes explicitly contrasted the laudatory illocutionary force of praising to the assertoric illocutionary force of asserting the truth of a proposition. Although in praising things one will invariably assert it has certain properties, the essential function or illocutionary force of praising cannot be reduced to assertoric force: “In Orations of Prayse, and in Invectives, . . . the designe is not truth, but to Honour or Dishonour.”<sup>106</sup> When we express honour we are not merely asserting something is the case.

This distinction between asserting and praising is crucial to Hobbes’s theology, because Hobbes proposed reductively to analyze all theological utterances that apparently attribute properties to God into speech acts whose *sole* function is to express honour or praise: “For in the Attributes which we give to God, we are not to consider the signification of Philosophicall Truth; but the signification of Pious Intention, to do him the greatest Honour we are able.”<sup>107</sup> Hobbes had made the same point in *Anti-White* by distinguishing between uttering *propositions*, which proclaim our *conception that* such-and-such is the case, and uttering *oblations*, which express our *desire to* honour something. Only the former are truth apt, but the meaning of apparently assertoric theological utterances is wholly determined by their expressive or oblativ character.<sup>108</sup>

<sup>105</sup> L 31.8: 560; 10.17: 136; 6.59: 96; see 15.40: 242. Praising is characterized as a species of worship at L 31.9: 562.

<sup>106</sup> L 8.6: 106. <sup>107</sup> L 31.33: 568. Holden (2015); Abizadeh (2017c). <sup>108</sup> AW 35.16.

Hobbes suggested, moreover, that praising something in evaluative terms is implicitly also to prescribe favouring it. Since normative utterances invariably signify speakers' passions, when praising actions as good, or saying one has a reason for action, speakers do not merely signify their *opinion* that it is good. They also signify their favourable conative attitude towards it, namely, their *desire* that it be favoured in some way. But according to Hobbes, one may express or signify one's passions in two different moods. Speech acts in the *indicative* mood are used by speakers to assert that something *is* the case, i.e., to assert a proposition.<sup>109</sup> By contrast, speech acts in the *imperative* mood are used to prescribe that something *be* the case, i.e., that the content of a proposition be realized.<sup>110</sup> Thus Hobbes distinguished between speech acts with assertoric illocutionary force (whose essential function is to assert the truth of a proposition) and those with prescriptive force (whose essential function is to tell someone what to do). He drew on this distinction in *Leviathan* to differentiate between speech acts that signify or express one's passions in the indicative mood in order to describe or *report* on them, and speech acts that signify or express one's passions in the imperative mood in order to *prescribe*:

The formes of Speech by which the Passions are expressed, are partly the same, and partly different from those, by which wee expresse our Thoughts. And first, generally all Passions may be expressed *Indicatively*; as, *I love, I feare, I joy* . . . but some of them have particular expressions by themselves, which neverthelesse are not affirmations, unlesse it be when they serve to make other inferences, besides that of the Passion they proceed from . . . The language of Desire, and Aversion, is *Imperative*; as, *Do this, Forbear that*; which when the party is obliged to do, or forbear, is *Command*; otherwise *Prayer*; or els *Counsell*.<sup>111</sup>

True, one does not prescribe anything to anyone if one signifies one's opinion of an object's goodness, and one's desire to favour it, in a purely *indicative* mood. Perhaps Hobbes thought this is the mood in which confessions of faith, in which one reports one's religious attitudes to others, are uttered (such as the Apostle's Creed "I believe in one God, the Father Almighty," etc.).<sup>112</sup> But as we have seen, the "analytical" or definitional interpretation of Hobbes's *evaluative* speech, according to which calling something "good" reduces to reporting on one's desire that it be favoured, is not viable. Simply to report one's beliefs and passions does not amount to praising: for Hobbes, to praise is always evaluatively to express one's

<sup>109</sup> L 2.10: 36; 4.11: 54; DC 3.6; 5.1.    <sup>110</sup> L 6.55: 94; 25.1: 398.    <sup>111</sup> L 6.55: 94. See DC 3.1.

<sup>112</sup> HNH 7/392, emphasis removed.

passions, at least implicitly, in the *imperative* mood. This is all the more true of directive normative speech: in saying that some attitude or action is a precept of reason, Hobbes was not merely making a prediction; he was prescribing the attitude or action.

We can see this more clearly by considering the role of normative speech in “counsel.” As the quoted passage indicates, to counsel others to assent to some proposition, hold some attitude, or undertake some action is to utter speech in an imperative mood and hence to prescribe. To be sure, to counsel others is also (1) indicatively to assert and argue that certain descriptive facts obtain and (2) to assert and argue that such facts provide a sufficient reason for some attitude or action. This is why “one may examine (when there is need) the truth, or probability of his [the counsellor’s] reasons, and of the grounds of the advise he gives.”<sup>113</sup> But to counsel others by telling them they have a reason is also, Hobbes asserted, (3) to express one’s desire that they assent to a proposition, hold an attitude, or undertake an action in telling them to do so in the imperative mood. In counselling others against being disposed to pride, one does not only (1) assert and argue that pride promotes war, and (2) assert and argue that this fact furnishes a reason for them to avoid pride; one also, on the basis of such assertions, (3) expresses one’s desire that they do so in telling them to avoid pride. Hobbes claimed “these words, *Sell all thou hast; give it to the poore; and follow me*, are Counsell; because the reason for which we are to do so, is drawn from our own benefit.”<sup>114</sup> That we gain from following Christ is the fact (1), which constitutes a reason to follow Christ (2), which is the basis for the imperative to follow him (3).

Reasoning-based reductionism accounts for the first two elements of Hobbesian counsel: it takes the first element at face value and reductively analyzes the assertion that those counselled have a decisive reason to avoid pride as the assertion that they would avoid pride upon reasoning rightly. But it has no account of the third element: Hobbes recognized that expressing one’s desire in the imperative mood cannot be reduced to an utterance in the indicative mood (such as that involved in reporting one’s desires). To be sure, descriptivists need not deny language can be used in the imperative mood to prescribe: they can acknowledge that one can command another to “Do this!” But they deny, against Hobbes, that prescriptivity is an essential feature of normative judgement and language. For descriptivists, telling people that  $\phi$ -ing is good, or that they have a reason to  $\phi$ , that they ought to  $\phi$ , or that  $\phi$ -ing is a dictate of reason

<sup>113</sup> L 25.15: 410.    <sup>114</sup> L 25.10: 402–404.

does not inherently imply “Do  $\phi$ !” It is merely to say they would  $\phi$  upon reasoning properly. By contrast, Hobbes took his own normative propositions to have been uttered in the imperative mood of a counsellor: when anyone (himself included) “shall offer unto us any other Rules, which the Sovereign Ruler hath not prescribed, they are but Counsell, and Advice.”<sup>115</sup> Thus when Hobbes asserted that natural law “prescribeth Equity,” he understood himself to be prescribing equity to his readers.<sup>116</sup> Hobbes was not *commanding* his readers to be equitable, but neither was he merely predicting they would be equitable upon reasoning properly.

Pure descriptivism, and hence the reasoning-based interpretation, is therefore incompatible with Hobbes’s ethics. Any adequate interpretation must account for the irreducibly laudatory and prescriptive illocutionary force of normative speech<sup>117</sup> – including Hobbes’s own.

This is precisely what motivates the noncognitivist, prescriptivist interpretation. On this reading, normative utterances are not truth apt: their essential function is to prescribe, not to describe or assert. The essentially prescriptive function of normative language reflects the essentially non-cognitive nature of normative judgements. They are neither erroneous nor veridical: they are not cognitive states that represent states of affairs or events as having certain properties, but are purely conative states that dispose agents to favour or disfavour states or events.<sup>118</sup> Although this abandons descriptivism, it remains a broadly naturalist interpretation. And it is still committed to the view that there are no irreducibly normative properties for normative judgements to represent or for normative speech to describe.<sup>119</sup>

This prescriptivist reading faces three decisive problems of its own, however. First, noncognitivism is belied by Hobbes’s intentionalism. As we have seen, even conative states such as desire are, for Hobbes, partly cognitive: they represent objects as having the property of goodness. It might be thought that noncognitivism could be salvaged by joining it to projectivism. The initial thought here is that scientific analysis would expose the projection, so that on reflection we may understand that a “good” object is no more than something disposing us to desire it and hence to see it as good. Our scientifically informed normative “judgement” that something is good would consist in the purely conative aspect of our

<sup>115</sup> L 42.43: 822.    <sup>116</sup> L 15.26: 236.    <sup>117</sup> This is Hare’s (1952) central objection to descriptivism.

<sup>118</sup> Holden (2016).

<sup>119</sup> Emotivists and prescriptivists often characterize their metaethical position as a rival to “ethical naturalism,” but they are specifically rejecting *descriptivism*. For the argument that expressivism can take the form of ethical naturalism, see Harman (1977).

desire – the urge to favour the object – plus our reflective cognitive awareness that the goodness we see is mere projection (rather than our unreflective representation of the object itself as good). Even this modified noncognitivist reading, however, is not viable. Hobbes insisted that evaluative language such as praise signifies normative judgements that do partly consist in representing the object as good: “Honour consisteth in the inward thought, and opinion of the Power, and Goodnesse of another,” and praise is “The forme of Speech whereby men signifie their opinion of the Goodnesse of any thing.”<sup>120</sup>

Second, pure prescriptivism is belied by Hobbes’s characterization of normative propositions as truth apt: people commit crimes “from defect in Reasoning,” sometimes because they presume “False Principles of Right and Wrong,” and sometimes because they draw “Erroneous Inferences from True Principles . . . in concluding, and resolving what to do.”<sup>121</sup>

Third, pure prescriptivism is belied by Hobbes’s characterization of natural laws as not only *prescriptive*, but *natural*. To be sure, because both descriptivists and prescriptivists can acknowledge imperatival thought and language, they can also acknowledge prescriptive facts (not reducible to mere descriptions or predictions). In particular, they can both acknowledge the existence of *artificial* rules – such as the rules of chess or of the Roman Catholic church – that constitute or regulate some human practice and prescribe to individuals what to do. Artificial rules are by definition *mind-dependent*: they exist just insofar as intentional agents are reflectively guided by them in representing them, from within their own internal, participant perspective, as rules. The prescriptivity of artificial rules is ultimately wholly a property of mental states and the speech acts expressing those rules. The rules of chess, for example, prescribe in the sense that the mental states in which participants represent them and the speech acts in which participants express them from an internal perspective are themselves prescriptive. (Their content includes: “Do this!”) Descriptivists and prescriptivists can therefore both give a perfectly naturalistic explanation of the genesis and nature of artificial rules and their prescriptivity, without reducing them to descriptions or predictions. The prescriptivity of the rules derives from and is constituted by the prescriptive character of the mental states and speech acts that constitute the practice. (One can also describe such rules from an external, observer perspective, but what one is describing is in part the prescriptive character,

<sup>120</sup> L 31.8: 560; 6.59: 96.      <sup>121</sup> L 27.10–12: 458–460.

from the participant perspective, of the intentions and speech acts constituting the rules.)

If the prescriptivity of rules must be explained by the imperatival character of the mental states or speech acts in which they are represented or expressed, however, then one cannot provide an account of any *natural*, mind-independent, epistemically objective prescriptive facts. Pure descriptivists seize the first horn of this dilemma: insofar as they take the *natural*, mind-independent character of natural law for granted, they deem its apparently *prescriptive* character to be illusory. Despite the apparently prescriptive surface grammar, when Hobbes articulated the laws of nature in speech he was just making predictions: natural laws are mind-independent, purely descriptive facts whose truth does not depend on being represented by anyone as true. Pure prescriptivists, by contrast, seize the second horn of the dilemma: insofar as they take the law of nature to be prescriptive, they must deny it is natural. It is artificial: its prescriptivity must derive from and be constituted by the imperatival character of the mental states or speech acts in which it is represented or expressed. The problem is that in calling natural laws *natural*, Hobbes was denying that they depend on mental representation or linguistic convention in this way. He explicitly rejected the view that natural laws are prescriptive only insofar as they are acknowledged by “the Consent of all Nations, or the wisest, and most Civill Nations,” or even “the Consent of all mankind.” A conventionalist “definition [of the law of nature] cannot be allowed,” because such laws are dictates of reason for all rational creatures, whether they acknowledge them or not.<sup>122</sup> A rule can be normative for those who *can* represent and follow it, even if they do not actually do so.

### 1.5 A Hybrid Theory: Reasoning-Based and Prescriptivist

Pure descriptivism fails as an interpretation of Hobbes because it misses the inherently laudatory and prescriptive illocutionary force of normative speech. By contrast, pure prescriptivism fails because it misses the truth-aptness of normative speech and its inherently assertoric illocutionary force. And neither interpretation accounts for naturally prescriptive facts.

These problems might therefore be resolved by a *hybrid* reading that combines a reasoning-based, descriptive component with a prescriptivist

<sup>122</sup> EL 15.1. Natural law is not in itself law in the proper, juridically obligatory sense; when Hobbes rejected as “absurd” Bramhall’s view that “the Law of Nature is a Law without our assent” (Q: 137/180), he was specifying what would make it properly law.

one. The hybrid reading claims that normative judgements are partly cognitive – insofar as they represent some object as having a certain naturalistic property, namely, that it would be desired upon reasoning rightly – and partly conative, insofar as they comprise a desire for the object. Concomitantly, it claims that normative propositions are always used both descriptively, to assert that upon proper reasoning some object would be desired or some action undertaken, and prescriptively: they signify a desire in an implicitly imperative mood and hence prescribe desiring the object or undertaking the action.

On the hybrid interpretation, to say that others have a reason to do something is not only to say they would do it upon reasoning rightly, but also to tell them to do it. More generally, to offer counsel is to (1) assert some proposition (as true), (2) predict the outcome of right reasoning in light of that truth, and (3) tell someone to assent to some proposition, adopt some attitude, or undertake some action. Thus to argue that “there is no Reason for any one, to devest himselfe of his” right of nature under insecure circumstances is to assert that divesting himself under such circumstances would court death, predict that people would not so expose themselves were they to reason properly, and tell everyone not to so expose themselves.<sup>123</sup> The first two elements are descriptive, but the third is irreducibly prescriptive. However, on this reading, when Hobbes used ‘reason’ as a count noun in the third, distinct sense to say that we have a reason to believe, desire, or do something, he did not mean we have an irreducibly normative reason to do so: he meant to predict that we would do so upon reasoning rightly and to tell us to do so.

The hybrid reading is therefore able to explain how natural laws can be both natural and prescriptive. They are natural insofar as the predictive fact they are used to assert is true independent of being acknowledged or asserted by anyone. And they are prescriptive in virtue of being used by intentional agents to prescribe dispositions or actions.

The hybrid interpretation is also able to explain Hobbes’s use of ‘reason’ as a mass noun in expressions such as ‘dictate of reason’. For Hobbes to say there is a dictate of reason against expressing “Hatred, or Contempt of another”<sup>124</sup> is to predict that one would, upon reasoning rightly, be disposed to avoid doing so and to prescribe that one be so disposed. By incorporating these two components, the hybrid account can explain

<sup>123</sup> Not having a reason to  $\phi$  does not entail having a reason not to  $\phi$ , but the context makes it clear Hobbes meant the latter as well.

<sup>124</sup> L 15.20: 234, italics removed.

why saying some action is against reason (mass noun) is equivalent to saying one has a reason (count noun) not to do it: in both cases, one is predicting and prescribing.

I conclude that the hybrid interpretation yields the most plausible interpretation of Hobbes as a naturalist who rejects irreducibly normative properties and facts. Yet even if this hybrid reading can explain Hobbes's use of 'reason' both as a count noun and as a mass noun, I shall argue in the [next chapter](#) that it ultimately fails because it is belied by Hobbes's use of 'reason' as a *verb*, i.e., by his account of *reasoning*. The hybrid interpretation fails properly to integrate the reasoning-based and prescriptivist elements that compose it: it fails to show how the prescriptivity of a reason is intrinsically *linked* to its status as the purported outcome of right reasoning. Furthermore, I shall argue in [Part III](#) that it also fails to account for Hobbes's use of 'obligation' in the juridical sense to denote reasons of the right. The only way to account for Hobbes's normative philosophy is to take him to have recognized the existence of reasons in the irreducibly normative sense. If Hobbes is a founder of modern naturalistic ethics, he founded a version that acknowledges irreducibly normative reasons.

*Mind, Action, and Reasoning*

Hobbes found the received view of human agency defended by late scholastics such as Suárez so nonsensical that he declared belief in it “rightly . . . numbred amongst the many sorts of Madnesse.”<sup>1</sup> On the “Mad” scholastic view, voluntary action arises from an incorporeal faculty, namely the will, which is free, self-determining, and capable of responding to practical reasons cognized and represented to it by an incorporeal intellectual faculty. External corporeal action (*exterior actus*) is voluntary just to the extent it is commanded (*imperati*) by the uncaused, rationally motivated, and intrinsically voluntary internal act of willing (*actus elicitus*).<sup>2</sup> Hobbes, by contrast, denied the existence of a non-corporeal soul and of a distinct, higher rational faculty of judgement, described mental phenomena in terms of matter in motion, and repudiated the freedom of the will. His hard-nosed mechanism inspired a deceptively simple, rival account of voluntary action: it is action resulting from the will, but the will is nothing more than the last passion in deliberation immediately prior to acting. There is no *actus elicitus* or voluntary act of willing.<sup>3</sup>

Hobbes’s readers have usually taken him also to have denied that the will and voluntary actions respond to judgements about normative reasons. On a widespread interpretation, Hobbes defended an entirely *passive* account of practical deliberation and willing, passive in the sense that they do not involve reflectively evaluating one’s desires, or endorsing and adopting them as one’s will on the basis of taking oneself to have a normative reason to do so. “Hobbes’s non-normative conception of deliberation,” as Terence Irwin has put it, has “no place” for a will that

<sup>1</sup> L 8.22: 122.      <sup>2</sup> Pink (2004, 2010).

<sup>3</sup> LN §20; Q: 236/300. For Hobbes’s materialism and determinism, see Brandt (1927); Mintz (1969); Lott (1982); Friedle (2012).

“results from the judgment that one option is better than the other, or from the judgment that we have stronger reasons for pursuing it.”<sup>4</sup>

On the dominant version of this passive interpretation, Hobbesian practical deliberation (and hence voluntary action) is unresponsive to (1) judgements about normative *reasons* because it is unresponsive to (2) the activity of *reasoning* through which reasons are discovered and reflected upon. Hobbesian practical deliberation is not responsive to reasoning and to cognitive judgements about reasons because either (a) deliberation wholly consists in the alternation of passions which, in turn, are purely conative mental states or urges (such that deliberation excludes cognitive mental phenomena altogether); (b) even if it includes cognitive mental phenomena, it excludes reasoning; or (c) even if it includes the cognitive process of reasoning, reasoning and judgements about reasons are impotent and so fail to help determine passions or actions.<sup>5</sup>

If this dominant version were correct, then it would rule out both the *hybrid* interpretation of Hobbes’s metaethics and the *irreducibly normative* interpretation defended in this book. First, recall that, according to the hybrid reading, Hobbes reductively analyzed normative speech acts into two components: each is an assertion about what actions agents would undertake (or what mental states they would have) upon reasoning rightly, joined to a prescription to undertake that action (or to be in that state). This clearly assumes rational agents can determine their beliefs, desires, and actions through *reasoning*. It purports to account for the difference between what people actually desire or do and what reason dictates they desire or do by equating the latter with what they would desire or do upon reasoning rightly. Second, the irreducibly normative reading of Hobbes would also be ruled out, because there can be no normative reasons for creatures incapable of responding to them. But according to the dominant passive interpretation, Hobbesian deliberation is unresponsive not only to *reasoning*, but also to cognitive judgements about *reasons*.

The upshot of the dominant interpretation is that Hobbes was a strong nihilist about affective and practical reasons: his talk of the dictates of reason and natural laws was either philosophically meaningless or false. This is to construe Hobbes’s account of passion, deliberation, and action as incompatible with any meaningful incorporation of dictates of reason or

<sup>4</sup> Irwin (2008: 105). See Riley (1982: 37–38, 40, 43); Sorell (1986: 95); Darwall (2000: 331, note 31); Pink (2011: 558).

<sup>5</sup> For (a), see Hampton (1986: 19); Paganini (2012: 327); Frost (2008: 101). For (b), see Hampton (1998: 320); Rutherford (2012: 207–208); Blau (2016). For (c), see Strauss (1963: 3, 81, 92).

laws of nature into his moral science. It is to read him as radically confused or incoherent.

In what follows, I begin by defending both the hybrid and genuinely normative readings against this strongly nihilist implication. I do so by showing that the dominant interpretation of Hobbesian practical deliberation as unresponsive to reasons and to reasoning misconstrues the historical significance of Hobbes's alternative to late-scholastic views about agency. I argue that for Hobbes (a) deliberation is a partly cognitive process; (b) amongst humans it normally involves reasoning; and (c) reasoning is not impotent, i.e., it can help to determine beliefs, passions, and therefore actions.

There is, however, an alternative interpretation of Hobbesian deliberation as passive capable of conceding all of this. According to this alternative, although practical deliberation can respond to *reasoning*, it nevertheless remains unresponsive to judgements about *reasons*. Although deliberation includes reasoning and reasoning is not impotent, reasoning is *itself* passive on this view, in the sense that, although it is a cognitive process, it is one in which agents merely experience a succession of mental states following purely associative mental structures. Reasoning is not an activity in which reasoners draw inferences from premises they endorse and thereby take to be reasons for belief or action: it is not responsive to judgements about reasons.<sup>6</sup>

This alternative interpretation is significant because while it is compatible with the hybrid reading, it rules out the irreducibly normative one. The hybrid reading requires only that passions and actions be responsive to *reasoning*; but the irreducibly normative reading requires that passions and actions be responsive to judgements about *reasons*. Partisans of the hybrid reading can therefore try to defeat the rival, irreducibly normative reading by appealing to this alternative version of the widespread view that Hobbesian deliberation is passive.

I end this chapter, however, by turning the tables on the hybrid reading. I argue that while the passive interpretation of Hobbesian reasoning is roughly accurate for non-linguistic reasoning, it fails as an interpretation of specifically *linguistic* reasoning: in linguistic reasoning, reasoners endorse or suppose propositions *from* which they inferentially reason.<sup>7</sup> To be sure,

<sup>6</sup> For interpretations according to which Hobbesian deliberation is passive either because deliberation excludes reasoning or because reasoning itself is passive – but which are unclear about which because they do not adequately distinguish between deliberation and reasoning – see Riley (1982: 34 ff.); Irwin (2008: 103–109).

<sup>7</sup> On acting “from” a reason or reason-responsively, see Skorupski (2010: 57–59, 462–464).

Hobbes repudiated incorporeal and self-determining mental faculties, but he did not deny humans' capacity for reflectively responding to and being guided by reasons. The historical interest of Hobbes's account lies precisely in his endeavor to provide a materialist, determinist account of reasoning, reason-responsive agency, and voluntary action not dependent on an incorporeal will or intellect. He appealed instead to the same faculty by which we perceive and imagine, albeit enhanced by the artifice of language. Hobbes pursued this agenda by characterizing responding to reasons as a causally determined activity in which agents form beliefs on the basis of rules for inference embedded in linguistic conventions.

I conclude that the hybrid reading fails because it cannot adequately account for Hobbes's conceptions of reasoning, deliberation, and counsel. Only an irreducibly normative reading can do so.

### 2.1 Practical Deliberation Is Partly Cognitive

The purely conative and non-reason-responsive conception of deliberation widely attributed to Hobbes is not without textual basis. His account of deliberation and voluntary action, given for the first time in *Elements*, is premised on a distinction between "Cognitive or Imaginative or Conceptive" mental states, on the one hand, and "motive" or conative mental states or passions, on the other.<sup>8</sup> Passions are conative states in the sense that they comprise an urge or disposition to act: they take the form of "appetite" or "desire" when they dispose us to favour some object, and "aversion" when they dispose us to disfavour it. *Elements* repeatedly describes these passions as "the beginning of all his voluntary motions" or "the Indeavour, or internall beginning of animall motion."<sup>9</sup> *Elements*' characterization of deliberation as a sequence of alternating passions for and against undertaking some action seems to suggest that the origin of voluntary action is purely conative: the "alternate succession of appetite, and feare, during all the tyme the action is in our power to doe, or not to doe is that we call DELIBERATION."<sup>10</sup> The "will," in turn, is merely "the last appetite" or aversion in deliberation, immediately prior to action, and the passions that precede and constitute the will are themselves involuntary: they cannot be willed into being via some *actus elicitus*. The will is the origin of all voluntary action, but is itself neither voluntary (in the sense of being willed) nor free (in the sense of being causally undetermined and/or self-determining).<sup>11</sup>

<sup>8</sup> EL 1.7.    <sup>9</sup> EL 5.14; 7.2.    <sup>10</sup> EL 12.1. Cf. AW 30.26; 37.4.

<sup>11</sup> EL 12.2–6. Cf. AW 30.27; LN; Q.

Moreover, deliberating and willing have no intrinsic relation to reasoning. As Hobbes made clear in *Leviathan*, although beasts do not reason in the strict sense, the “alternate Succession of Appetites, Aversions, Hopes and Feares, is no lesse in other living Creatures than in Man; and therefore Beasts also Deliberate,” and “Beasts that have *Deliberation*, must necessarily also have *Will*.” Hobbes contrasted his account of the will, as “*the last Appetite in Deliberating*,” to the absurd view propounded “by the Schooles” according to which the will “is a *Rationall Appetite*” and hence a faculty only of creatures endowed with reason.<sup>12</sup> One does not voluntarily choose which desire to act on: a desire results in voluntary action because the internal motion or “endeavour” constituting its material basis outweighs any contrary aversions to result in external bodily motion.<sup>13</sup> No wonder so many have concluded that Hobbesian deliberation is purely conative, excludes reasoning, and is unresponsive to reasoning and to judgements about reasons.

But this conclusion can be sustained only on the basis of a selective, tendentious presentation of the textual evidence. First, Hobbes was clear that conative states, and hence the will, are caused by *cognitive* mental states such as perception or opinion. As he put it in *Elements*, “externall objects cause Conceptions, and Conceptions appetite.” This is partly why “our wills follow our opinions, as our actions follow our wills” or, as he put it *Leviathan*, “the Actions of men proceed from their Opinions.”<sup>14</sup> Indeed, what distinguishes “*Vitall*” from “*Voluntary motion*” is that the former originates unconsciously, as when blood circulates, while the latter depends on the mind’s *cognitive* power of thought or imagination. The “small beginnings of Motion, within the body of Man,” which “are commonly called ENDEAVOUR [LL: *Conatus*]” are for voluntary actions to be found in the faculty of conception or imagination: because “Voluntary motions, depend alwayes upon a precedent thought . . . it is evident that the Imagination is the first internall beginning of all Voluntary Motion.”<sup>15</sup>

Second, our conceptions have an at least causal role in deliberation: “in Deliberation, the Appetites, and Aversions are raised by foresight of the good and evill consequences, and sequels of the action whereof we Deliberate” – where foresight is a species of thought or imagination.<sup>16</sup> According to *Leviathan*’s definition of deliberation,

When in the mind of man, [1] Appetites and Aversions, Hopes and Feares, concerning one and the same thing, arise alternately; and [2] divers good

<sup>12</sup> L 6.51–53; 92. Cf. AW 30.28; 37.4.      <sup>13</sup> Cf. DCv 13.16.

<sup>14</sup> EL 12.1; 12.6; L 18.9; 272. See DC 25.9.      <sup>15</sup> L 6.1; 78; LL 6: 79.      <sup>16</sup> L 6.57; 94; 3:7: 42.

and evill consequences of the doing, or omitting the thing propounded, come successively into our thoughts; so that sometimes we have an Appetite to it; sometimes an Aversion from it; sometimes Hope to be able to do it; sometimes Despaire, or Feare to attempt it; the whole summe of Desires, Aversions, Hopes and Fears, continued till the thing be either done, or thought impossible, is that we call DELIBERATION.<sup>17</sup>

This definition asserts that the successive “thoughts” (of consequences) play a causal role in giving rise to the passions involved in deliberation. The passage also assumes that such a “TRAYNE of Thoughts” – or “*Mentall Discourse*” – occurs during (or at least intermittently in between episodes of) deliberation.<sup>18</sup> The question is whether Hobbes thought that such mental discourse plays a *merely* causal role or whether it can also form a constitutive part of deliberation.<sup>19</sup>

We have considerable evidence that Hobbes thought of deliberation as partly constituted, and not just causally influenced, by the train of thoughts. His most extensive discussion of deliberation occurs in *Of Liberty and Necessity*, composed in 1645 in his debate with John Bramhall, the Anglican bishop who had urged Hobbes’s acquiescence to the late-scholastic view of agency and free will.<sup>20</sup> Hobbes there repeatedly described deliberation in constitutively cognitive terms, asserting that voluntary action “presupposes some precedent deliberation, that is to say, some consideration and meditation of what is likely to follow,” and that “deliberation is the consideration of the good and evil sequels of an action to come.”<sup>21</sup> Nor was this some aberrant view with which Hobbes only temporarily flirted in 1645. Hobbes offered a cognitive characterization in *De Corpore* as well: “seeing living creatures have sometimes appetite and sometimes aversion to the same thing, as they think [*putant*] it will either be pleasing or hurtful to them; while that vicissitude of appetites and aversions remains in them, they have that series of thoughts [*cogitationum*] which is called *deliberation*.”<sup>22</sup> He reaffirmed the view in *The Questions Concerning Liberty, Necessity, and Chance* of 1656, insisting “I do indeed conceive that deliberation is an Act of *Imagination* or *Fancie*.”<sup>23</sup>

It is striking that *Of Liberty and Necessity* seems to *equate* this cognitive characterization of deliberation with the conative one he had given in *Elements* and would give again in *Leviathan*: “deliberation is nothing but

<sup>17</sup> L 6.49: 90.

<sup>18</sup> Counselling consists in the “expression of those Conceptions which cause in us the expectation of Good, while we deliberate.” EL 13.5.

<sup>19</sup> For the claim that mental discourse is constitutive of deliberation, see van Apeldoorn (2012).

<sup>20</sup> For the context, see Jackson (2007). <sup>21</sup> LN §8; §33. <sup>22</sup> DC 25.13. <sup>23</sup> Q: 309/401.

alternate imagination of good and evil sequels of an action, or, which is the same thing, alternate hope and fear or alternate appetite to do or quit the action of which he deliberates.”<sup>24</sup> Bramhall attacked this equation as a clear equivocation: “Sometimes he makes it [deliberation] to be a consideration, or an act of the understanding, sometimes an imagination, or an act of the fancy, sometimes he makes it to be an alternation of passions, hope and fear . . . So he makes it I know not what.”<sup>25</sup> Hobbes directly responded to this in *Questions*, and to appreciate his response we must recall a key feature of his philosophy of mind.

Hobbes distinguished two types of cognitive mental state: *sensory perception* and *imagination*. Recall that sensory perception, which lies at the origin of all mental states,<sup>26</sup> is produced when an “Externall Body, or Object” presently exerts pressure on one’s sensory organ and thereby causes internal bodily motions – giving rise to a “*Representation or Apparence*” of the object. Imagination, in turn, arises when such internal motions persist despite there being no external pressure on one’s sensory organs from the object represented – as in memory, when “after the object is removed, or the eye shut, wee still retain an image of the thing seen,” or fiction, “as when from the sight of a man at one time, and of a horse at another, we conceive in our mind a Centaure.”<sup>27</sup>

Although Hobbes usually used the terms ‘image’ or ‘imagination’ narrowly for the latter type of mental state, yet since sensory perceptions are representations of intentional objects too, he also sometimes used ‘imagination’ more broadly to include sensory perception.<sup>28</sup> Indeed, ‘imagination’ for Hobbes could be used in two further ways – one even narrower than his usual usage, the other even wider than the broad usage just indicated. On the one hand, given its etymology, Hobbes suggested that in its narrowest and most literal sense ‘imagination’ means *visual* representation, and is used for other types of representation only metaphorically – the more literal general terms being ‘representation’ and ‘appearance’.<sup>29</sup> On the other hand, Hobbes used ‘imagination’ in the widest, most inclusive sense to comprise not just sensory perceptions and imaginations, but the passions as well. It is this widest usage that is of particular interest here.

What justifies this widest usage is that Hobbes took the passions not only to be *caused* by cognitive mental states, but themselves to *be* a species of cognitive state. It is true that Hobbes distinguished between conceptions

<sup>24</sup> LN §26.   <sup>25</sup> DTL: 206.   <sup>26</sup> L 1.2: 22.   <sup>27</sup> L 1.1: 22; 1.4: 22; 2.2: 26; 2.4: 28.

<sup>28</sup> E.g. L 45.1: 1012; 46.16: 1078. See AW 30.4; Schuhmann (2004: 20).   <sup>29</sup> L 2.2: 26.

(in a narrow sense) and passions – a distinction illustrated by his remark that “The formes of Speech by which the Passions are expressed, are partly the same, and partly different from those, by which we express our Thoughts.”<sup>30</sup> Whereas the passions are conative states, thoughts in the narrow sense are not. But on Hobbes’s account these two types of mental state share an essential feature: they are both cognitive. Bramhall failed to realize Hobbes was an *intentionalist* who took all mental states to be intentional states representing some object. All mental states are *directed* on an intentional object they are about, and they represent that object in a particular or *aspectual* way, as having some specific properties.<sup>31</sup> As Hobbes put it in *Leviathan*, “besides Sense, and Thoughts [LL: *Imaginationem*], and the Trayne of thoughts, the mind of man has no other motion,” and “the Thoughts [LL: *Cogitationes*] of man . . . are every one of them a *Representation* or *Apparence*, of some quality, or other Accident of . . . an *Object*.”<sup>32</sup> Thus visually to perceive something is to see it as having a certain colour and shape; aurally to perceive something is to hear it as having a certain (e.g. loud, high-pitched) sound; and so on. If passions are mental states, if mental states are all conceptions, and if conceptions are all representations, then passions must themselves be conceptions of an intentional object and with cognitive, representational content.

That Hobbes took passions to be a species of conception in the widest sense is not merely a matter of inference: he said so explicitly on numerous occasions. In *Leviathan*, he characterized “desire, or other passion” as a type of “Passionate Thought,” and asserted that “all our affections are but conceptions.”<sup>33</sup> Indeed, Hobbes *defined* desire as a type of thought: “This Endeavour, when it is toward something which causes it, is called APPE-TITE, or DESIRE,” where “This Endeavour” Hobbes had already identified with “a precedent thought.”<sup>34</sup> That passions are a type of thought is precisely what Hobbes asserted in *Questions* in response to Bramhall’s charge that his account of deliberation is equivocal. Bramhall had failed to realize “that consideration, understanding, reason, and all the passions of the mind; are imaginations . . . when we hope or fear, we imagine things good or hurtful to our selves.”<sup>35</sup> Hobbes’s claim that all mental states are “imaginations” is to be understood in the widest sense of the term: Hobbes meant that all mental states, including passions, consist in an “apparence” representing some object as having certain properties.

<sup>30</sup> L 6.55: 94.    <sup>31</sup> Crane (2001).    <sup>32</sup> L 3.11: 46; I.1: 22; LL 3: 47; I: 23.

<sup>33</sup> L 3.3: 38; 4.24: 62. See EL 7.4; DC 25.12.    <sup>34</sup> L 6.1–2: 78.    <sup>35</sup> Q: 278–279/358–359.

Passions arise when the internal bodily motions giving rise to perception or imagination are towards or away from the cognitive mental state's intentional object. The consequent conative mental *state* is called delight or pleasure when the motions are towards the intentional object, and trouble, molestation, or displeasure when away from it: "the apparence, or sense of that motion," i.e., the mental representation to which the motion gives rise, "is that wee either call DELIGHT, or TROUBLE OF MIND."<sup>36</sup> The motions themselves – i.e., the *material basis* of the mental state – are called "desire" or "appetite" in the former case, and "aversion" in the latter. Finally, the *content* of such a mental state, i.e., the aspect under which the intentional object of pleasure or displeasure is represented, is as having the property of being good or evil:

*Pleasure* therefore, (or *Delight*,) is the apparence, or sense of Good; and *Molestation* or *Displeasure*, the apparence, or sense of Evill. And consequently all Appetite, Desire, and Love, is accompanied with some Delight more or lesse; and all Hatred, and Aversion, with more or lesse Displeasure and Offence.<sup>37</sup>

The pleasure accompanying "all Appetite" arises from *conceiving* the object of one's desire *as good*.<sup>38</sup> In *De Corpore*, Hobbes asserted that because pleasure and displeasure are intentional, cognitive mental states, they can even be thought of as "another kind of sense" alongside sensory perception, "namely, the sense of pleasure and pain [*sensio voluptatis et doloris*]."<sup>39</sup>

A passion therefore has two aspects: a cognitive mental representation of an object *as* a good or evil object, and a conative urge or disposition – an endeavour or *conatus* – to favour or disfavour the object. Hobbesian passions are therefore not purely functional states or urges: they are hybrid states joining together a conception of something as good or evil to the impulsion to favour or disfavour it.<sup>40</sup> Hobbes had intimated this hybrid view in his *Objectiones* to Descartes, suggesting that a passion consists in a cognitive thought joined with an impulsion to which the thought gives rise: "For what is fear of a charging lion if not the idea of a charging lion plus the effect which this idea produces in the heart, which in turn induces in the frightened man that animal motion which we call 'flight'?"<sup>41</sup> The evaluative component of the conception, which lies at the genesis of passion, becomes a constitutive part of the passion itself.

That the passions are intrinsically cognitive, evaluative mental states is why although sometimes Hobbes characterized voluntary action as

<sup>36</sup> L 6.9–10: 82.    <sup>37</sup> L 6.11: 82. See Darwall (2000).    <sup>38</sup> L 6.10: 82.    <sup>39</sup> DC 25,12; cf. 2.15.

<sup>40</sup> Tuck (1996: 183–185); Schneewind (1998); cf. Scanlon (1998: 37–41).

<sup>41</sup> O 6.

immediately preceded by a passion, at other times he characterized the moment prior to action in cognitive terms, as “the last dictate of the judgement, concerning the good or bad that may follow on any action.”<sup>42</sup> This judgement becomes a constituent element of the last passion immediately prior to action. Hobbes took “the will necessarily to follow the last dictate of the understanding,” which simply means “that the will follows the last opinion or judgment immediately preceding the action, concerning whether it be good to do it or not.” Because one’s evaluative judgement about the consequences both *causes* and becomes a *constituent part* of one’s final conative state prior to action,<sup>43</sup> Hobbes could immediately restate the claim that a person’s voluntary action “necessarily follows the present thought he has of the good or evil consequence” or “follows his opinion of the goodness or harm of it” as the apparently equivalent claim that “those actions that follow immediately the last appetite are voluntary.”<sup>44</sup> The judgement dictating action arises from representing reasons to oneself and *weighing* them: “Medea saw many reasons to forbear killing her children, yet the last dictate of her judgment was that the present revenge . . . outweighed them all, and thereupon the wicked action necessarily followed.”<sup>45</sup> Hobbes’s characterization in *De Cive* – that “deliberation is nothing else but a weighing [*ponderatio*], as it were in scales [*in bilance*], the conveniences, and inconveniences of the fact we are attempting”<sup>46</sup> – must be understood in these partly cognitive terms.

This cognitive conception of passion, voluntary action, and deliberation leaves the door open to the possibility that reasoning – itself a cognitive process – could both help determine how an agent acts and be a constituent part of deliberation. This is because voluntary actions result from passions, passions are caused by beliefs and constituted by judgements, and reasoning is a process by which rational agents can form beliefs.

Hobbes suggested that as a matter of fact reasoning often does play a constitutive role in deliberation. In *Elements*, he wrote of “those Conceptions which cause in us the expectation of Good, while we deliberate,” and called the articulation of the consequences of potential actions the “arguments whereupon” the agent can “deliberate within himselfe.”<sup>47</sup> In *Leviathan*, he asserted that since “in Deliberation, the Appetites, and Aversions are raised by foresight of the good and evil consequences,” the agent who “Deliberates best himselfe” is one who

<sup>42</sup> LN §11.    <sup>43</sup> LN §23.    <sup>44</sup> LN §25.    <sup>45</sup> LN §23; cf. Q: 39/52.    <sup>46</sup> DCv 13.16.

<sup>47</sup> EL 13.5.

“hath by Experience, or Reason, the greatest and surest prospect of Consequences.”<sup>48</sup> His most explicit statement comes in *Questions*, where he asserted that “the reason which a man useth in deliberation” is “the same thing that is called Deliberation” – not because Hobbes was asserting that “deliberation is always (as it is not) rational,” but because deliberation *can* partly consist in reasoning.<sup>49</sup> Indeed, in creatures capable of reasoning, it usually does.

## 2.2 Is Hobbesian Reasoning Passive?

We must therefore reject the view that Hobbesian deliberation is (a) purely conative or (b) excludes reasoning. We must also reject the view that (c) Hobbesian reasoning is impotent. Cognitive states and reasoning can determine desire and action, and reasoning is sometimes a constitutive element of deliberation. Hobbes’s conception of deliberation therefore provides no grounds for a strongly nihilist reading of his metaethics: it is compatible with both the hybrid and the irreducibly normative readings.

Partisans of the hybrid reading, however, might claim that Hobbes’s conception of *reasoning* excludes the irreducibly normative reading. True, reasoning and cognitive states can determine action and desire, and reasoning itself is a cognitive process. But perhaps Hobbesian deliberation is nevertheless a passive process, i.e., one that does not involve reflectively evaluating one’s passions and coming to form one’s will on the basis of taking oneself to have a reason to do so. Perhaps deliberation is passive because (d) reasoning itself is a purely passive, albeit cognitive, process. Perhaps reasoning merely comprises a succession of mental states that follow innate associative structures, rather than mental states one reflectively endorses. Reasoning would not, on this interpretation, involve reasoning *from* considerations agents take to be normative *reasons*. If this were right, then even if deliberation were to include reasoning, it would nevertheless remain passive, and agents’ actions, although responsive to reasoning, would remain unresponsive to judgements about reasons.<sup>50</sup> This would provide decisive grounds for rejecting the irreducibly normative interpretation in favour of the hybrid one.

One way to understand such a reading is to distinguish between four, progressively stronger senses in which mental phenomena might be said to

<sup>48</sup> L 6.57: 94; see L R&C.1: 1132.

<sup>49</sup> Q: 279/359–360; 284/365, ‘rational’ missing in 1656 edition, italics removed. See AW 30.28; 38.4.

<sup>50</sup> See note 6.

be cognitive. (1) In the weakest sense, to call mental phenomena cognitive is to say they are or involve intentional states with *representational content*: for example, a conception of an object O as having property F. (2) In a somewhat stronger sense, to call mental phenomena cognitive is to say they are or involve intentional states with *propositional content* that is *truth apt*: for example, a conception whose content is the proposition *p* that object O has property F. (3) In an even stronger sense, to call mental phenomena cognitive is to say they are or involve intentional states with truth-apt propositional contents to which the agent *internally assents*. Assenting to a proposition *p* does not amount to merely *having* a conception with the propositional content *p*; one could, for example, have a conception with content *p* where one's relation to that content – one's propositional attitude or intentional mode – is to dissent from or merely to suppose the proposition. Finally, (4) mental phenomena might be called cognitive in the strongest sense insofar as they consist in taking some fact to furnish a normative reason to endorse and internally assent to some proposition *p*, or in endorsing and internally assenting to a proposition *on the basis of* taking some fact(s) to furnish a sufficient reason to do so. Call these the *representational*, *propositional*, *propositional-attitude*, and *reason-responsive* senses of cognition. The irreducibly normative interpretation takes Hobbesian reasoning to be cognitive in all four senses.

(1) The evidence that Hobbes deemed mental phenomena cognitive in the first, weakest sense consists in his intentionalism, according which *all* mental states have representational content. (2) Moreover, Hobbes thought that the content of any merely representational, non-propositional mental state could in principle be transposed and articulated in propositional form: the conception *of* an intentional object O *as* having property F can be transformed into a conception *that* object O has F. We can put this by saying that, according to Hobbes, any merely representational state is potentially propositional, or that it has *implicit* propositional content.<sup>51</sup> But making the implicit propositional content of our conceptions explicit marks an important cognitive difference for Hobbes, as we shall see, and such a transformation is what we effect (and can effect only) when we “transferre our Mentall Discourse, into Verbal” discourse, by joining subject and predicate to form sentences via the copula.<sup>52</sup>

<sup>51</sup> Hobbes marked implicit propositional content via the term *considerare*: to consider an intentional object is to conceive it *as* having distinct aspects and to focus on them. AW 2.1; L 4.14–18: 58–60; Nuchelmans (1983: 131–132). See O 6.

<sup>52</sup> L 4.3: 50; DC 5.9.

The passive interpretation acknowledges that reasoning is cognitive for Hobbes in the first two senses; in a strict version it denies, although in a moderate version it grants, that reasoning is cognitive in the third, propositional-attitude sense; and in any version it denies that mental phenomena can be cognitive in the fourth, reason-responsive sense. Mental life is therefore limited to conceiving of objects as having certain properties or to conceiving that they have certain properties; it possibly extends to assenting to such intentional contents; but it involves no mental act of judging that we have a reason to assent to a proposition, or of endorsing (the truth of) a proposition on the basis of such a judgement. Reasoning therefore could not consist in reflectively responding to (what the reasoner takes to be) normative *reasons*. Rather, it involves a succession of conceptions we are caused to experience in virtue of purely associative mental structures and not in virtue of being guided by rules or precepts: we do not directly *reason from* considerations we take (or suppose) to be reasons.

The strongest evidence in favour of this passive reading is Hobbes's cognitive monism: his rejection of a higher intellectual faculty distinct from the imagination.<sup>53</sup> According to Hobbes, agents can reflect on their own mental states, but such reflection merely involves *remembering* those states via the imagination. Even "the judgement" by which we observe differences between our conceptions is a species of the imagination, and does not involve a distinct, higher faculty.<sup>54</sup> Judging the differences between two conceptions – say, between a visual and an aural sensory perception of the same object – simply involves conceiving their differences by jointly remembering them in a single conception. As Hobbes put it in *Anti-White*, "*comparatio* is a mental discourse consisting of three acts [*actibus*] (i.e., three *imaginatioes* continued among themselves), of which the first is of one thing, the second of another, and the third an *imaginatio* of the difference between the other two."<sup>55</sup> Hobbes was not using 'actibus' here as the name of genuine mental *operations*: it is the name of occurrent mental states, namely imaginations, arising in sequence. Here is Hobbes's response, in *De Corpore*, to scholastics who insisted that a higher intellectual faculty is needed to explain how one can reflect on or compare conceptions: "you will say, by what sense shall we take notice of sense? I answer, by sense itself, namely, by the memory which for some time remains in us of things sensible, though they themselves pass away.

<sup>53</sup> O 5; DC 5.9.    <sup>54</sup> On "the judgement" in this sense, see L 8.3; 104–106 ff.

<sup>55</sup> AW 30.14; see 30.32.

For he that perceives that he hath perceived, remembers.”<sup>56</sup> Even reasoning is a species of imagination for Hobbes.<sup>57</sup> A partisan of the strict passive reading therefore concludes that for Hobbes, reflecting on one’s conceptions cannot amount to, and reasoning cannot involve, judging and/or internally *assenting* to the veracity of the (implicit) propositional content of one’s conceptions.

A difficulty for the strict passive reading is that Hobbes clearly contrasted the mental states of perceiving or imagining something to the mental state of *believing* the implicit propositional content of the perception or imagination.<sup>58</sup> It is one thing to perceive or imagine the sun in the water, another to believe the sun is really in the water.<sup>59</sup> Furthermore, with respect to propositional contents, he explicitly distinguished between merely “conceding” or affirming a proposition, and genuinely “receiving” or internally assenting to it. As he put it in *De Cive*, “we grant [*Concedimus*] Propositions sometimes which notwithstanding we receive not into our mindes [*in animum non recipimus*],” such as when we provisionally suppose (*supponere*) a proposition in order to reason with it, outwardly profess (*profiteri*) or confess (*confiteri*) it out of “feare of the Lawes,” or simply concede (*simpliciter concedere*) it out of civility to others. Indeed, Hobbes distinguished between different modes of internal assent (*assensu interno*) or “inward perswasion of the minde [*animi persuasio interna*],” contrasting the propositional attitude of knowing (*Scire*) to those of thinking or opining (*opinari*) and of having faith (*Fides*).<sup>60</sup>

Confronted with this textual evidence, partisans of the passive reading might simply retreat to the moderate version: concede that (3) Hobbes distinguished between different intentional modes and held a notion of assenting to propositions, but insist that (4) he had no notion of assenting to propositions on the basis of taking some fact to furnish a reason to do so. Alternatively, one might defend the strict passive reading by insisting that Hobbes merely sounded as if he had a notion of assenting to a proposition, but did not genuinely have one. The issue is how to explain the structural difference, in the case of representational states, between a perception and a belief with identical content and, in the case of propositional states, between the different types of attitude one might have towards the same

<sup>56</sup> DC 25.1. Leijenhorst (2002: 82–83). <sup>57</sup> Q: 278–279/358–359.

<sup>58</sup> *Leviathan* contrasts two senses of ‘belief’. In the first, where one’s mental state is based on faith in another person, “Beleefe” combines “two opinions; one of the saying of the man; the other of his vertue.” In the second sense, it concerns only the proposition’s content. L 7.5: 100. I am using ‘belief’ in Hobbes’s second sense.

<sup>59</sup> DC 5.1. See O 5. <sup>60</sup> DCv 18.4. Cf. EL 6.2–7; O 13.

propositional content. The explanation cannot appeal to a distinct intellectual faculty that stands above the imagination and judges which conceptions to assent to.

The strict passive reading concludes that Hobbes did not appeal to any genuine structural difference or mental act of assenting at all: states such as belief or opinion, which Hobbes characterized as a kind of internal assent, arise from an alternation of perceptions and/or imaginations in the same way that, on the purely passive account of deliberation, the will arises from an alternation of desires and appetites. The claim is that Hobbes provided a *reductive analysis* of assent in terms of experiencing a mental state: belief or opinion is nothing more than the *last* conception one has in a mental discourse about the matter. Consider the close parallel *Leviathan* draws between practical deliberation and the discursive process of forming a judgement:

that which is alternate Appetite, in Deliberating concerning Good and Evil; the same is alternate Opinion, in the Enquiry of the truth of *Past*, and *Future*. And as the last Appetite in Deliberation, is called the *Will*; so the last Opinion in search of the truth of Past, and Future, is called the JUDGEMENT, or *Resolute* and *Finall Sentence* of him that *discourseth*. And as the whole chain of Appetites alternate, in the question of Good, or Bad, is called *Deliberation*; so the whole chain of Opinions alternate, in the question of True, or False, is called DOUBT.<sup>61</sup>

The passage seems to characterize mental discourse as a process in which agents passively experience a succession of alternating conceptions (“Opinions”): the agent is, as it were, alternately in the grips of various conceptions, and settles on a “judgement” when the alternation of conceptions runs its course. An agent’s “Opinion” just seems to be whatever conception has gripped him at any given point in time: to perceive the sun in the water is, at least momentarily, to hold the opinion that it is in the water. One “judges” or “believes” that the sun is in the water only in the thinned out, reductive sense that the last conception one has in discoursing about the matter is a conception of the sun being in the water. On this account, there is nothing that *structurally* differentiates “Opinion” or “judgement” from other perceptual or imaginative conceptions. It is as if Hobbesian reasoners were thoroughgoing sceptics who suspend cognitive judgement over the conceptions they experience during mental discourse – not because they are committed sceptics, but because, as Irwin suggests, they are incapable of doing otherwise.<sup>62</sup>

<sup>61</sup> L 7.2: 98.      <sup>62</sup> Irwin (2008: 33–34).

It might be objected that this passive interpretation goes against Hobbes's characterization of reasoning as the mental *operation* or act of *adding* or subtracting the constituent properties or the antecedent causes (both of which Hobbes called "causes"<sup>63</sup>) that we conceive the objects of our reasoning to have. As Hobbes put it in *De Corpore*,

By RATIOCINATION, I mean *computation* . . . *Ratiocination*, therefore, is the same with *addition* and *subtraction* . . . So that all ratiocination is comprehended in these two operations of the mind, addition and subtraction.<sup>64</sup>

Similarly in *Leviathan*:

When a man *Reasoneth*, hee does nothing els but conceive a summe totall, from *Addition* of parcels; or conceive a Remainder, from *Subtraction* of one summe from another: which (if it be done by Words,) is conceiving of the consequence of the names of all the parts, to the name of the whole; or from the names of the whole and one part, to the name of the other part.<sup>65</sup>

But Hobbes's notion of mental calculation here is perfectly compatible with the interpretation according to which we passively experience a series of conceptions without reflecting on and being guided by a rule.

We can see this by focussing on non-linguistic reasoning. It is true that Hobbes characterized the mental calculations involved in reasoning as typically carried out in linguistically explicit, propositional form. Indeed, prior to 1643, Hobbes held that language is necessary for reasoning, and some of Hobbes's commentators have thought he held this view throughout his career.<sup>66</sup> But as the phrase "if it be done by Words" indicates, this is mistaken: in his mature works Hobbes allowed for non-linguistic reasoning in which "a man reckons without the use of words."<sup>67</sup> In *De Corpore* he explicitly addressed the question of "how by the *ratiocination* of the mind, we add and subtract in our silent thoughts, without the use of words" by illustrating how reasoning could proceed prior to language. Even when "no appellation had yet been given to anything," if "a man see something afar off and obscurely," he will "have the same idea of that thing for which now, by imposing a name on it, we call it body," i.e., he will have a conception of it *as a body*. When, subsequently, he sees it "now in one place and now in another, he will have a new idea thereof" *as animated*. Thirdly, when he

<sup>63</sup> Engel (1962: 332–333); de Jong (1986: 136–138); Malcolm (2002: 153); Pettit (2008: 22–23).

<sup>64</sup> DC 1.2. <sup>65</sup> L 5.1: 64. On reasoning as calculation, see Losonsky (2001: chapter 3).

<sup>66</sup> E.g. Pettit (2008: 42–43). For Hobbes's early view, which rules out non-linguistic reasoning, see EL 5.11; O 4; AW 4.1; 30.22; 30.28. Pécharman (1992); Jessep (1999) recognize non-linguistic reasoning. Cf. Losonsky (2001: 49–52).

<sup>67</sup> L 5.5: 68. This passage undermines Gert's (2001) distinction between reasoning and reckoning.

observes the object more closely and sees in it “signs of a rational mind,” he will form a conception of it *as rational*. The moment of calculation finally arrives when, “by looking fully and distinctly upon it, he conceives all that he has seen as one thing,” such that he forms a new “idea . . . compounded of his former ideas, which are put together in the mind in the same order in which these three single names, *body*, *animated*, *rational*, are in speech compounded into this one name, *body-animated-rational*, or *man*.” That is to say, he forms a single, structured conception of the object as corporeal, animated, and rational.<sup>68</sup> The so-called mental “operation” simply consists in having a mental *state* of a particular (structured) kind.

This description of purely mental, non-linguistic reasoning seems to confirm the strict passive interpretation. Reasoning here seems to amount to no more than a sequence of conceptions of an intentional object as having several distinct properties, followed by a conception of those conceptions as having the same intentional object, and, finally, a compound conception of the object as having all of the several properties combined. There is no indication of the agent internally assenting to the implicit propositional content of his conceptions and consequently taking those conceptions to provide him with reasons *from which* he draws a conclusion. The agent simply *has* those conceptions in sequence.

This passive reading is not completely off the mark: with one important qualification to be noted shortly, it provides an accurate interpretation of Hobbes’s account of purely mental, non-linguistic reasoning. It is, however, wholly inadequate as an interpretation of *linguistically articulated* reasoning, in which “we turn the reckoning of the consequences of things imagined in the mind, into a reckoning of the consequences of Appellations.”<sup>69</sup>

### 2.3 Linguistic Reasoning is Reflectively Reason-Responsive

Linguistic reasoning differs from non-linguistic reasoning in several important respects according to Hobbes.<sup>70</sup> (1) The first and perhaps most decisive difference is that linguistic reasoning is, but non-linguistic reasoning is not, *propositional*. Whereas there can be linguistically articulated mental propositions, there are no purely mental, non-linguistic propositions. Propositions are *essentially* linguistic: a proposition by definition is “a

<sup>68</sup> DC 1.3. <sup>69</sup> L 4.9: 52.

<sup>70</sup> For Hobbes’s philosophy of language, see Hungerland and Vick (1981); Zarka (1999); Pettit (2008); Abizadeh (2015).

speech [*oratio*] consisting of two names copulated.”<sup>71</sup> There can be no purely mental proposition because the implicit propositional content of intentional states can be made explicit only linguistically.

On the one hand, we *cannot* make the implicit propositional content of intentional states explicit without language because, according to Hobbes’s nominalist account of mental states, all intentional states are representations of a *particular* object. We may very well have a conception of the hot celestial body we call the sun. But we cannot have *universal* conceptions, such as a conception of hot celestial body.<sup>72</sup> Nor can we have purely *abstract* conceptions, such as a conception of the property of heat. We can conceive of a particular object *as* having some property – we can conceive the sun as hot – but we cannot have a conception of any property independently of the object whose property it is. For on Hobbes’s nominalist account, properties are not themselves entities, but the features or “accidents” of objects. Yet to make explicit, without language, the propositional content implicit to our conception would require forming a purely mental proposition consisting of, first, a conception of the intentional object (the proposition’s grammatical subject) and, second, a conception of the object’s property independently of the object itself (the proposition’s predicate) as if the property could exist separately – which Hobbes took to be impossible.

On the other hand, we *can* make propositional contents explicit linguistically because language does not merely consist in the semantic imposition of *names* on objects; it also consists in a syntactic structure by which names are “ordered in speech,” joined together as subject and predicate.<sup>73</sup> “SPEECH . . . [consists] of *Names* or *Appellations*, and their *Connexion*.”<sup>74</sup> The syntactic structure of speech, which serves “to shew the Consequence, or Repugnance of one name to another,” is usually expressed by using some word as a copula, “as wee use the Verbe *Is*” in English, although Hobbes acknowledged that languages could also form propositions “without any Verb answerable to *Est*, or *Is*, or *Bee*,” because placing “two names in order may serve to signifie their Consequence.”<sup>75</sup> The copula, whether its function is performed via a syncategorematic word or sentential syntax, is

<sup>71</sup> DC 3.2, italics removed. On Hobbes’s treatment of the proposition, see Nuchelmans (1983: chapter 7); Zarka (1999: chapter 3).

<sup>72</sup> EL 5.6; L 4.6: 52; DC 5.8; 6.15. At DC 6.15, Hobbes wrote of *notionem universalem*, but immediately clarified that the term is for the image of a *particular* item. Nor does language make universal conceptions possible, *pace* Pettit (2008).

<sup>73</sup> DC 2.5.      <sup>74</sup> L 4.1: 48.

<sup>75</sup> L 46.16–17: 1078–1080. On the copula and proposition, see EL 5.9; DC 3.2.

supposed to explain the unity of the proposition, i.e., to explain how a proposition such as ‘Socrates is a man’ differs from a mere list such as ‘Socrates, being, man’.<sup>76</sup> According to Hobbes, the verb ‘to be’ or its syntactic equivalent unifies subject and predicate in virtue of enabling the proposition as a whole to signify our conception of two conceptions as being conceptions of the same intentional object, namely, the conceptions that would be signified by the predicate and by the subject, respectively. When we utter ‘Socrates is a man’, for example, we signify our conception that our conception of Socrates and our conception of some rational, animate body are conceptions of the same intentional object: we do not, as with a list, conceive of Socrates and conceive of a man in two separate conceptions; we signify a single, structured conception of some intentional object, in which we consider the object as possessing both properties.<sup>77</sup>

(2) The second difference between non-linguistic and linguistic reasoning arises from the impossibility of universal conceptions. “When a man reckons without the use of words,” he can reason only about the particular objects that can be conceived and hence may reason only “in particular things.”<sup>78</sup> But once people have introduced “Universall” names, which are the names “imposed on many things, for their similitude in some quality, or other accident,”<sup>79</sup> they can form, and reason with, *universal propositions*.<sup>80</sup>

One might wonder how creatures capable of neither universal nor abstract conceptions could invent universal names. The answer is that, for Hobbes, although all conceptions are of particular objects, they are intrinsically *aspectual* (we conceive objects as having some particular properties) and *comparative* (we inherently conceive objects as being similar to or different from other conceptions’ intentional objects). To have a conception of an intentional object just is to conceive it *as* similar to or different from other objects: no faculty beyond imagination is needed. Hobbes argued that without comparing the representational content of our current conception to previous ones in memory, we would not even have sensory perceptions:

For by sense, we commonly understand the judgment we make of objects by their phantasms; namely, by comparing and distinguishing those

<sup>76</sup> See Gibson (2004). Geach (1962: 60) erroneously asserts that Hobbes held “that the copula was superfluous” and did not distinguish propositions from mere lists. Hobbes merely conceded that the function of the copula can be performed syntactically. See Ott (2004: 36).

<sup>77</sup> DC 5.9. Nuchelmans (1983: 132–133). <sup>78</sup> L 5.5: 68. <sup>79</sup> L 4.7: 52. <sup>80</sup> DC 4.8; cf. 3.5.

phantasms; which we could never do, if that motion in the organ, by which the phantasm is made, did not remain there for some time, and make the same phantasm return. Wherefore sense . . . hath necessarily some memory adhering to it, by which former and later phantasms may be compared together, and distinguished from another.<sup>81</sup>

Language is therefore not necessary for comparing the properties of intentional objects; language simply enables us to do explicitly, in propositional form, what we already did in representational form via consideration.<sup>82</sup> It is this inherently comparative nature of conception that enables the formation of universal names (on the basis of similarities and differences we conceive in sundry intentional objects).

(3) The third, related difference<sup>83</sup> between non-linguistic and linguistic reasoning arises from the impossibility of abstract conceptions. Since we cannot conceive a property apart from the object of which it is the property, in non-linguistic reasoning we cannot *treat* or “consider” a property separately from the object of which it is the property.<sup>84</sup> Once reasoning is linguistic and hence propositional, however, we can separate subject from predicate, and so can invent *abstract names* for properties and treat them separately from the concrete name of the object of which they are the properties. (Concrete names are names of objects, and abstract names “are the names of the accidents and properties” by which one object “is distinguished from another” – e.g. the word ‘heat’.<sup>85</sup>)

Because there can be no abstract conceptions, Hobbes argued humans were incapable of inventing abstract names for properties until they formed linguistic propositions: it is only via a linguistic proposition, with its distinct subject and predicate, that one can make the propositional content of one’s intentional states explicit and treat the abstract name of the property independently of the concrete name of the object whose property it is.<sup>86</sup> This confers on linguistic reasoning a tremendous advantage; indeed, “ratiocinating . . . cannot for the most part be undertaken” without abstract names.<sup>87</sup> Yet abstract names also harbour danger: the ability to treat abstract *names* separately from concrete names might lead some, as Hobbes accused Aristotle and his followers of having done, to think that

<sup>81</sup> DC 25.5. <sup>82</sup> Pettit (2008).

<sup>83</sup> Hobbes illustrated the second and third differences in L 4.9: 52–54. Cf. DC 6.11.

<sup>84</sup> Hobbes’s phrase for what is treatable in discourse or reasoning is “whatsoever can enter into, or be considered in an account.” To treat something is to “bring [it] into account, [or] consider” it. L 4.14–18: 58–60. See HNH: 8/394; Jessep (1999: 76–80).

<sup>85</sup> L 4.16: 58. See LL Appendix 1.75: 1175–1177.

<sup>86</sup> DC 3.4. See DC 2.13; AW 27.1; Nuchelmans (1983: 135–137). <sup>87</sup> DC 3.4.

one could conceive of *properties* separately from the objects whose properties they are, and consequently to think that properties are entities with independent existence. This “doctrine of *Separated Essences*” Hobbes ridiculed as nothing but an absurd and politically nefarious abuse of language.<sup>88</sup> People confuse the ability to treat names separately with the ability to consider properties independently because they fail to realize that abstract names in isolation do not mark or signify anything at all: they contribute to the power of a *proposition* to mark or signify, and the conception signified by a proposition containing an abstract name can be nothing but a conception of some particular intentional object considered in some way. Thus to say that abstract names are the names of properties is just to say they are the names of concrete things insofar as we consider them in a particular way.<sup>89</sup>

One might wonder how creatures incapable of abstract conceptions could separate subject from predicate to form a proposition in the first place. The answer is that the predicates of the first propositions did not contain abstract names: they contained universal names more general than the name contained in the proposition’s subject – for example, ‘Socrates is a man’. On Hobbes’s account, humans could form universal names, propositions, and consequently abstract names thanks to the inherently comparative nature of conception. No higher faculty is required.

(4) The fourth difference is that, because language enables us to render a conception’s propositional content explicit, in the form of a linguistic proposition, it also enables us to conceive the proposition and hence explicitly treat its content as the object of judgement. Without language, we can conceive of the sun *as* hot, and we can conceive of our conception of the sun. But with language we can conceive *that* the sun is hot and conceive of the proposition articulated linguistically in our initial conception; we can consequently explicitly assent to the linguistically articulated propositional content of our conceptions. Upon explicitly (linguistically) articulating propositional contents, we are able to treat the proposition as an object of thought, reasoning, and judgement, and thereby assent to (or dissent from or suppose) its truth, because we can conceive of the proposition *as* true – indeed, we can have a conception whose propositional content is that the proposition is true. Although “some brute creatures, looking upon the image of a man in a mirror, may be affected with it, as if it were the man himself, and for this reason fear it or fawn upon it in vain; yet they do not apprehend it as true or false, but only as similar; and in this

<sup>88</sup> L 46.18: 1082; 46.15: 1076–78; DC 3.4.    <sup>89</sup> LL Appendix 1.77: 1177.

they are not deceived.”<sup>90</sup> By contrast, when language users *assent to* or *dissent from* the proposition that the image of a man in a mirror is itself a man, they conceive of the proposition ‘*spectrum est homo*’ as true or false.<sup>91</sup> Beasts, like sceptics, merely yield to their conceptions, but language users can judge and reflectively assent to propositional contents.

Hobbes marked the difference between implicit and explicit propositional contents by distinguishing between veridicality and error in general, and truth and falsity in particular. A mental representation of a table’s surface *as* square-shaped may be veridical or erroneous. But a mental representation with merely implicit propositional content cannot, according to Hobbes, be true or false: only the proposition *that* the table’s surface is square-shaped can be true or false. Truth is a property of propositions, not of things or of mental states: conceptions can be truth apt only if they are linguistically articulated: “these words *true*, *truth*, and *true proposition*, are equivalent to one another; for truth consists in speech, and not in the things spoken of.”<sup>92</sup> Hence humans are subject to two different types of error: errors of sensory perception (*sentiendo*, *sensu*) and of non-linguistic (*tacita*) thought (*cogitatione*), on the one hand, and errors of affirmation or proposition, on the other. Only the latter “error deserves the name of *falsity*, as arising not from sense, nor from the things themselves, but from pronouncing rashly.”<sup>93</sup> Errors of the former type “are common to all things that have sense,” while the latter “have no place but amongst such living creatures as use speech.”<sup>94</sup>

A critic might grant that language enables us (a) to make conceptions’ propositional content explicit, and thereby enables us (b) to treat propositions as an object of thought, but might nevertheless object that it remains unclear, on Hobbes’s account, how language could enable us (c) to *assent to* the propositions language enables us to conceive, if prior to language we could not assent to our conceptions. It is, after all, one thing to hold an explicitly articulated proposition in view and another to adopt a particular propositional attitude towards it. The answer is that, for Hobbes, just as pre-linguistic intentional states have implicit propositional content, which can be made explicit linguistically, so too are pre-linguistic agents able implicitly to have distinct intentional modes towards the representational content of their conceptions. In other words, even non-linguistic reasoning

<sup>90</sup> DC 3.8; cf. 5.1. <sup>91</sup> DC 3.7. <sup>92</sup> DC 3.7. Cf. L 4.11: 54.

<sup>93</sup> DC 5.1. Hobbes’s repetition of ‘*tacita*’ at DC 6.11 shows that he meant *non-linguistic* (rather than just *silent*, as per the English translation).

<sup>94</sup> DC 5.1; 3.8. Compare with the Cartesian distinction between formal and material truth (Nuchelmans 1983: 50–51).

does not fully accord with the strict passive interpretation (according to which Hobbes did not genuinely differentiate between different intentional modes such as merely perceiving and assenting). Hence the necessary qualification, announced earlier, to the strict passive account of purely mental reasoning.

We can see all this in Hobbes's account of perspective. The question Hobbes was addressing in [chapter 3](#) of [part II](#) of *A Minute or First Draught of the Optiques* (1646) is how a drawing in a two-dimensional plane surface can come to represent a three-dimensional object for an observer. Hobbes answered that the illusion of a three-dimensional object is created when seeing the two-dimensional figure triggers the *memory* of a three-dimensional object like the one the drawing represents, so that we end up conceiving the three-dimensional object of memory rather than the two-dimensional object of vision before us – or, more precisely, we end up conceiving the two-dimensional object of vision before us as having the properties we conceive in the three-dimensional object of our memory:

the Reason why those figures generally have an apparence [different] from their figures made in the playne of perspective is this. That when we have in memory the Originalls which they are made to represent, the plaine itselfe is not (to speake properly) seene, butt the Originall remembred . . . when wee behold a perspective and acknowledge nott anything it represents butt itselfe, then is the fancie of the beholder, vision, namely the vision of the plaine, Butt when wee conceyve by it a Gallery, Landskip or other thing represented by it, then is the fancy of the beholder to bee called memorie, though that memorie bee raised and confirmed by the lines drawne on the plaine.<sup>95</sup>

True, the plane is not seen in the “proper” or literal sense that we see it as a two-dimensional plane. But to experience the illusion of perspective one must form a compound conception arising from simultaneously seeing the two-dimensional drawing and remembering the three-dimensional object it resembles. If one did not see the two-dimensional drawing at all – if memory were not “raised and confirmed” by what one visually *perceives* – then one would be merely imagining a three-dimensional object, rather than imagining the two-dimensional drawing before one's eyes *as* three-dimensional.<sup>96</sup> Those who remember having seen drawings will be aware they are not actually seeing a three-dimensional object: they will recognize that their conception is a compound of sensory perception and memory. This is the moment of implicit dissent from or assent to

<sup>95</sup> MDO II.3.9, accepting Malcolm's (2002: 226) emendation.    <sup>96</sup> Malcolm (2002: 227).

(“acknowledge,” as Hobbes put it) the implicit propositional content of one’s conception. Such acknowledgement is implicit in that one passively *experiences* the content of one’s mental state as veridical or illusory. This experience is not something intrinsic to the intentional content, but is an attitude agents implicitly hold towards that content. As Hobbes put it in *De Corpore*, in non-propositional errors of perception or imagination, “as when, by seeing the image of the sun in water, we imagine the sun itself to be there,” or “when from any sign we vainly imagine something to be signified, which is not,” the “deception proceeds neither from our senses, nor from the things we perceive; but from ourselves while we feign such things as are but mere images to be something more than images.”<sup>97</sup> If we take the content of our conception to signify the presence of an object with certain objective, mind-independent properties, we do so only implicitly, not in the form of explicit propositions. We cannot explicitly *adopt* or endorse such an attitude until language allows us to make the implicit propositional content explicit.

This undermines the strict passive interpretation of reasoning because this notion of explicitly assenting cannot be reductively analyzed into merely *having* or experiencing a conception with representational content. At best, a partisan of the strict passive reading must claim that, although Hobbes thought he had a notion of structurally distinct intentional modes, unbeknownst to him his cognitive monism ruled out any such notion. Yet even that claim would be unmotivated: since whenever we conceive something we must conceive it in some particular way, there is no reason why intentional modes must be explained by a faculty distinct from the faculty of mental representation.

A more promising objection might be that, in the case of intentional modes that are specifically propositional attitudes, Hobbes’s account proves too much. The worry here is not that his account fails to accommodate the notion of assenting to a proposition, but that it fails to accommodate any other propositional attitude, because it cannot distinguish *forming* a proposition from *assenting* to it. The implication would be that, on Hobbes’s account, one cannot but assent to a proposition whenever one unites subject and predicate. Philosophers since John Stuart Mill have often accused their early-modern predecessors of conflating the unity of the proposition and assenting to its content in this way,<sup>98</sup> and Hobbes might be thought to be a typical culprit insofar as he equated ‘proposition’ with ‘affirmation’ (or ‘negation’).<sup>99</sup> But Hobbes need not be saddled with

<sup>97</sup> DC 5.1.    <sup>98</sup> SLRI I.5.1. Cf. Ott (2004: 41–49).    <sup>99</sup> EL 5.9. DC 3.6.

such an obvious mistake. Affirming does not, for Hobbes, carry assertoric force: he explicitly distinguished between merely affirming or “conceding” a proposition and internally assenting to or “receiving” it. To affirm a proposition in Hobbes’s sense is merely to *form* a proposition via the copula, that is, to connect a subject to predicate rather than produce a mere list.<sup>100</sup> It remains a separate question whether one assents to it.

We must therefore reject the strict passive interpretation of Hobbesian reasoning: Hobbes distinguished between different propositional attitudes such as assenting to and supposing propositional contents, and these attitudes play a constitutive role in reasoning. Partisans of the passive interpretation must retreat to the moderate version, which denies only that Hobbesian reasoning includes assenting on the basis of taking some fact to be a reason. As I shall now argue, even this moderate interpretation fails.

(5) That linguistic reasoning is propositional also explains its fifth distinctive feature: it is *inferential*. Non-linguistic reasoning consists in a series of conceptions the agent is caused to have in virtue of innate associative mental structures – for example, structures in virtue of which one is caused to conceive of two previous conceptions as being of the same object and, subsequently, to form a compound conception of the object that combines all the properties separately represented in the previous conceptions. This is calculation (addition), and it ends with conceiving the “consequences” of one’s train of thoughts, but it is not inference: one is not guided by any rule.<sup>101</sup> Hobbes assimilated “consequences” here to temporal *succession*, not logical derivations.<sup>102</sup> Linguistic reasoning, by contrast, consists in drawing inferences: it is governed by “the Method of Reasoning,” which amounts to the constitutive rules of logic reasoners apply to *derive* consequences.<sup>103</sup> Linguistic reasoning involves not merely *assenting to* (or supposing) propositions, but doing so *on the basis of* having already assented to (or supposed) prior propositions: it involves *reasoning from* the content of premises treated as a reason to assent to (or suppose) some further proposition.

The upshot is that the passive reading of Hobbes, whether in its strict or moderate version, is unstable: what enables mental phenomena to be cognitive in the second sense, i.e., propositional and truth apt, is precisely what for Hobbes enables them to be cognitive in the strongest

<sup>100</sup> An “*affirmative* proposition” is merely one “whose predicate is a positive name.” DC 3.6.

<sup>101</sup> *Contra* Losonsky (2001: 47, 50). <sup>102</sup> Pécharman (1992). <sup>103</sup> L 46.11: 1060; DC 6.

sense, i.e., involving assent to some proposition on the basis of taking some fact(s) to furnish a sufficient reason for doing so.

(6) We can reiterate this point by articulating a sixth distinguishing feature of linguistic reasoning. While in the former one's sequence of thoughts is governed by innate associative mental structures, in the latter it can proceed via *artificial* relations governed by rules of language such as syntax. Language enables humans to transform, via propositions and inferences, the innate patterns of mental discourse. True, Hobbes rejected the view that reasoning could be uncaused, or that one could experience a conception at will. But none of this rules out influencing one's train of thought via assenting, judging, and calculating. Humans can use language to affect the sequence of their thoughts, not because they can escape from the determinist chain of causes but because their succession of thoughts can follow patterns governed by rules they themselves have previously established by convention.

Hobbes's description of the origin of language highlights its transformative role. Hobbes argued that language originally arises privately, when, as he put in *Elements*, an "arbitrarily imposed" vocal sound is used by someone "for a marke to bringe to his minde, some Conception, concerning the thinge on which it is imposed."<sup>104</sup> The motivation for using names as marks is that mental discourse amongst non-linguistic creatures suffers from an important "defect": it follows purely innate patterns of association. Without language, in "the Succession of Conceptions in the minde . . . one Conception followeth not another according to our Election, and the neede we have of them, but as it chanceth us to heare, or see, such things as shall bring them to our minde."<sup>105</sup> Without language, a particular conception comes to mind because an external object stimulates one's sensory organs, leading to sensory perception, or because one has just had a conception that one's mind is disposed to associate with another conception, thanks to previous temporal conjunctions. One will imagine rain only after perceiving or imagining something else, such as a thick cloud, that one has previously perceived or imagined immediately before or after rain. One's sequence of thoughts may even be guided or "*regulated* by some desire":<sup>106</sup> one might take events observed in regular temporal succession to be signs of each other's presence, but also to stand in a means-end relation, such that when one imagines a desired end, one

<sup>104</sup> EL 5.2. See L 4.3: 50; DC 2.3; 6.II. Hungerland and Vick (1981) deny the private origin of Hobbesian language; for refutations, see Ross (1987); Abizadeh (2015).

<sup>105</sup> EL 5.1. See EL 4.2; L 3.2: 38. <sup>106</sup> L 3.4: 40.

might be disposed, in regulated mental discourse, to think of its presumed means. Thus if seeing thick clouds causes one to imagine rain, then the desire to avoid getting wet, which involves representing to oneself staying dry during rain, might prompt one to imagine being in a cave, which one has previously experienced in conjunction with staying dry. This is an example of prudential thinking, of which beasts are also capable. But both unregulated and regulated trains of non-linguistic thought operate strictly on the basis of innate *associationist* laws of psychology.<sup>107</sup>

On the associationist principles just noted, if each time upon perceiving or imagining rain one utters the sound “rain,” then one will be causally disposed to imagine rain simply by uttering “rain” in the future. Yet merely imposing marks in this way would be insufficient for transforming one’s sequence of thoughts, because one could not cause oneself to utter “rain” or think of the word ‘rain’ if nothing in the external world has already caused one to perceive rain or triggered one’s memory of it.<sup>108</sup> Hobbes understood this: he did not think one could recall, at will, individual conceptions in *isolation*. Rather, he was trying to show how we could affect the *sequence* or “succession of conceptions” one after another. According to Hobbes, the genesis of language involves not just the invention of names *qua* marks but also the invention of rule-governed ways to connect names together into sentences or propositions: not just using vocal sounds semantically as marks but syntactically combining names into sentences to affirm relationships between names – paradigmatically “by the helpe of this little verb IS.”<sup>109</sup>

It is therefore the invention of *propositions*, not just names, that transforms our thought sequences, by enabling us to have sequences that do not solely follow past *associations*. Once names are combined to form propositions and draw inferences, one’s trains of thought are governed not merely by *empirical* relations such as constant conjunction but also by the extended activity of reasoning made possible by the *rule-governed* relations embedded in linguistic syntax: “the Invention of names” and ways to connect them enable humans to call “to their remembrance the necessary coherence of one Conception to another” in a chain of “RATIOCINATION or Reasoning.”<sup>110</sup> Hobbes’s point was not that we could now conjure up whatever thought we would at will but that thought sequences could now be ordered according to the linguistic rules we ourselves have established by artifice – and hence ultimately by our own will or “Election.”

<sup>107</sup> L 3.3-9: 38–44.   <sup>108</sup> Martinich (2005: 141–142).

<sup>109</sup> EL 5.9-10. In this I disagree with Nuchelmans (1983: 125).   <sup>110</sup> EL 5.13, 5.11.

This is why Hobbes's mature formulation in *Leviathan* defines marks as devices not for recalling particular conceptions in isolation but for "the Registring of the Consequences of our Thoughts" – i.e., for recording the results of our *ratiocination*. "Speech serveth to the remembrance" of *conclusions*, not merely through the imposition of names but by "the imposing of *Names*, and the *Connexion* of them."<sup>111</sup> Names serve as mnemonic devices, but language also requires logical devices for ordering names.<sup>112</sup> This is why, although pre-linguistic humans may reason in the sense of calculating, only language users are "capable of Inferring, Concluding, and of all kind of Reasoning."<sup>113</sup>

(7) This points to a seventh difference between non-linguistic and linguistic reasoning, which arises from the fact that, without language, agents have very limited *memory*. Without language, one can remember isolated experiences, and one's train of memories can even be regulated by desire, but it is practically impossible to keep track of memories in the orderly manner required for extended reasoning: one could not reason "beyond a syllogism or two, by reason of the weakness of memory."<sup>114</sup> Language makes long chains of reasoning possible and, once equipped with universal and abstract names, enables one to draw conclusions about entire classes of objects and articulate and remember them as general theorems. A reasoner "can by words reduce the consequences he findes to generall Rules, called *Theoremes*, or *Aphorismes*," so that "the consequence found in one particular, comes to be registred and remembered, as an Universall rule . . . [which] delivers us from all labour of the mind, saving the first."<sup>115</sup> Language makes specifically philosophical reasoning possible, and some "general Rules" reasoning discovers are "for the direction of . . . actions."<sup>116</sup>

In sum, language transforms reasoning in at least seven ways. It enables: propositions; universal propositions; abstract names treated separately from concrete names; explicitly assenting to or dissenting from propositions; inferences; causing one's train of thought to follow artificial linguistic rules rather than innate associative structures; and enhanced memory.<sup>117</sup> On Hobbes's account, a higher faculty is required neither to conceive, judge, and assent to the propositional content of our conceptions nor to draw inferences from propositions. As long as our thoughts are

<sup>111</sup> L 4.3: 50; 4.5: 52. Although *Elements* does not *define* marks in this way, it makes the same point: "the Invention of names" works "by calling to their remembrance the necessary coherence of one Conception to another." EL 5.13. Cf. DC 2.1.

<sup>112</sup> Engel (1959: vi, 77, 85). <sup>113</sup> L 46.17: 1080. <sup>114</sup> DC 6.11. See L 4.10: 54.

<sup>115</sup> L 5.6: 68; 4.9: 54. See DC 6.11. <sup>116</sup> Q: 141/186.

<sup>117</sup> These seven differences are not exhaustive. See AW 2.1.

linguistically articulated and their propositional contents made explicit, the imagination is sufficient to conceive of propositions and the reason-relations they express. The capacities required for active, reason-responsive reasoning turn out to be latent within the selfsame faculty of conception or imagination: “besides Sense, and Thoughts, and the Trayne of thoughts, the mind of man has no other motion; though by the help of Speech, and Method, the same Facultyes may be improved to such a height, as to distinguish men from all other living Creatures.”<sup>118</sup> Hobbes’s rejection of a higher intellectual faculty provides no support for the passive interpretation.

#### 2.4 The Failure of the Hybrid Reading

Hobbesian linguistic reasoning is not a process in which conceptions succeed one another in virtue of purely associationist psychological laws. It is, rather, a *rule-governed* activity of inferentially drawing conclusions from premises by operating on their explicit, linguistically articulated propositional contents (adding and subtracting) and taking them to be normative reasons. This is a strongly cognitive account of reasoning. In rejecting the scholastic account, Hobbes was not excluding the possibility of responding to reasoning or to reasons. He was offering an account of human agency and reflectively responding to reasons compatible with determinism.

Hobbes’s account of reasoning therefore provides evidence against the hybrid, rather than irreducibly normative, interpretation. Recall that on the hybrid reading, to say that someone has a reason to  $\phi$  is just (a) to predict that he would  $\phi$  upon reasoning rightly, and (b) to tell him to  $\phi$  in order to cause him to  $\phi$ . Concomitantly, to take *oneself* to have a reason to believe something is just (a) to predict that one would believe it upon reasoning rightly and (b) to tell oneself to believe it in order to cause oneself to believe it. But this account implies that when agents engage in reasoning, they are simply trying to predict the outcome of reasoning correctly and to cause themselves to believe that outcome by telling themselves to. They could not *reason from* premises to conclusions, as Hobbes believed rational agents do when they reason linguistically: they could not take any consideration to count as a normative reason for inferring any conclusions. This is a mistaken interpretation: on Hobbes’s account, reasoners are not just telling themselves to adopt some attitude or undertake some action

<sup>118</sup> L 3.11: 46.

they have in mind prior to reasoning; they are trying to *discover* the reasons they have for inferring conclusions they may not antecedently have in view. Insofar as rational agents are making reasoned predictions about the outcome of reasoning, they do so *on the basis* of reasons. Their prediction does not constitute what a reason *is*: a law of nature is “a Precept, or generall Rule, found out by Reason.”<sup>119</sup> What excited Hobbes so much about Euclid’s geometrical reasoning was precisely this inferential ability to draw conclusions.<sup>120</sup>

A parallel issue arises in the intersubjective context of giving advice. Recall that for Hobbes, to counsel others is (1) to assert and argue in the indicative mood that certain descriptive facts obtain, (2) to assert and argue in the indicative mood that such facts provide a sufficient reason to  $\phi$ , and (3) to prescribe in the imperative mood that one’s advisees  $\phi$ . The hybrid interpretation accounts for these elements in the following way: (1) it takes the first element at face value (the counsellor asserts some descriptive facts); (2) it analyzes the second element as a prediction that the advisees would  $\phi$ , upon reasoning rightly in light of these facts; and (3) it interprets the third element as the counsellor expressing his desire that the advisees  $\phi$  by telling them to do so. Hence, unlike purely descriptivist or prescriptivist readings, this hybrid reading can separately account for each of the three elements of counselling. What it cannot do is adequately account for the intrinsic *relation* between the first two elements and the final one.

On the hybrid reading, when a counsellor predicts to others that they will  $\phi$  upon reasoning properly, and tells them to  $\phi$ , he fails to articulate any irreducibly normative reasons to  $\phi$  as he predicts and prescribes; indeed, he fails to articulate any irreducibly normative epistemic reasons to believe the prediction. Whatever evidence a counsellor cites for the prediction can at best merely *cause* them to believe it. There is therefore no intrinsic connection between the counsellor’s prescription and the factual claims he makes.

On Hobbes’s account, by contrast, there is supposed to be an intrinsic link between a counsellor’s affirmation that some fact  $p$  furnishes a reason to believe, desire, or do something and his prescription that one believe, desire, or do it. Those who *command* presume their say-so is a sufficient reason for action. By contrast, the pretension of those who *counsel* – whatever they actually believe – is that their advisee has a reason independent of their say-so: “COUNSELL, is where a man saith, *Doe*, or *Doe not this*, and deduceth his reasons from the benefit that arriveth by it to him to whom he

<sup>119</sup> L 14.3: 198. <sup>120</sup> BL I: 332; Jesseph (1999: 5).

saith it.”<sup>121</sup> True counsellors purport to tell one to  $\phi$  precisely *on the basis* of  $p$ : the fact is supposed to furnish not only the counsellor’s motivational reason for prescribing that one  $\phi$  but also the normative reason for one to  $\phi$ . This is why “one may examine (when there is need) the truth, or probability of his [the counsellor’s] reasons, and of the grounds of the advise he gives.”<sup>122</sup>

For counsellors to sever this intrinsic link between advice and reasons is to become *corrupt* rather than *good* counsellors. Consider counsellors who argue that pardoning some criminal is in the interest of sovereign and commonwealth. If they vehemently *exhort* their sovereign, they are “corrupt Counsellours” who “in their speeches . . . make use of Similitudes, Metaphors, Examples, and other tooles of Oratory, to perswade” him. They take themselves to be merely trying, as a result of their vehement desire that their sovereign believe and do as they say, to cause the sovereign to believe the proposition and to act accordingly – regardless of the proposition’s truth. A “good Counsellour,” by contrast, takes himself to be trying to cause his sovereign to believe the proposition and act accordingly by offering epistemic reasons to believe it insofar as it is true. This is precisely why he relies solely on “the rigour of true reasoning.”<sup>123</sup>

The hybrid reading corresponds to *corrupt* counsel, not good counsel: the prescription of corrupt counsellors merely expresses their strong desire that their advisee do or believe as they say; it is not grounded in any reasons their advisee has for doing so. Even if counsellors desire that their advisee do as they counsel, it is not their own desires per se, but their advisee’s interests that are supposed to be the normative *basis* for their advice: it is a defining feature of counsel that “he that giveth Counsell, pretendeth onely (whatsoever he intendeth) the good of him, to whom he giveth.”<sup>124</sup> The object of a good counsellor’s desire in giving counsel, and the normative basis for the particular counsel he gives, is his advisee’s good. There is an intrinsic link between a good counsellor’s affirmation that some fact  $p$  furnishes a reason to  $\phi$  and his prescription that the advisee  $\phi$ .

Here is how Hobbes put matters in *Elements*: “if reasoning aright I winne not Consent (which may very easily happen) from them that being confident of their owne knowledge weigh not what is said, the fault is not mine but theirs. For as it is my part to show my reasons, so it is theirs to bringe attention.”<sup>125</sup> The “fault” of those who do not heed Hobbes’s lessons is irreducibly normative and cannot be explained in terms of

<sup>121</sup> L 25.1–2: 398. Cf. EL 13.5–6; 29.4; DCv 14.1.      <sup>122</sup> L 25.15: 410.      <sup>123</sup> L 25.6–9: 400–402.

<sup>124</sup> L 25.3: 398.      <sup>125</sup> EL 1.3.

a conditional prediction; when Hobbes wrote of “my part,” he was writing of the normative responsibility rational agents have to reason rightly. Thus when Hobbes appealed to natural law to criticize others’ dispositions and actions, he was not merely predicting what they would (be disposed to) do under descriptively specifiable circumstances and expressing his desire that they do it in telling them to do it. When in *Leviathan* he accused the Foole of defending a position “against the reason of his preservation,”<sup>126</sup> he was not solely (a) claiming that upon reasoning rightly the Foole would come to agree with Hobbes and (b) telling the Foole to do as he said. He was also (c) arguing that the Foole has a normative reason to refrain from his proposed course of action. If the Foole had responded to Hobbes by granting the prediction that, upon reasoning rightly, he would refrain from violating covenants, but then went on to wonder whether he presently had any normative reason to refrain from violating covenants, he would have misunderstood Hobbes’s argument.

I conclude that the hybrid interpretation of Hobbes’s metaethics is not viable: Hobbes was committed to the notion of reasons in the genuine and irreducibly normative sense of the term.

<sup>126</sup> L 15.5: 224.



PART II

*Reasons of the Good*

That Hobbes classified ethics as a natural science does not itself imply that all normative propositions are reducible to descriptive claims about naturalistic facts. Even physics is grounded, in Hobbes's view, in a *philosophia prima* concerning the definitions of the foundational, most basic terms of science (such as 'body' and 'motion').<sup>1</sup> A true first philosophy might show that some of these terms and principles, on which all the sciences are erected, are themselves normative. Thus geometry and physics apparently involve claims about reasons to believe. The true challenge posed by Hobbes's classification of ethics is to show why it consists primarily, albeit not exclusively, in descriptive claims about mental phenomena. The challenge concerns not the nature of Hobbes's metaethical views and his conception of what a reason or goodness *is*, but of his views in normative ethics about what reasons and goods *there are*.

Most who take this challenge seriously locate the ground of normativity in the psychology of agents themselves – most typically in conative mental states. Thus many commentators attribute to Hobbes a desire-based theory of practical *reasons*, according to which individuals have genuinely normative reasons for action only in virtue of the fact that the action would promote the satisfaction of their present desires. This is often combined with attributing to him a substantively subjectivist, desire-based theory of *value*, according to which what *makes* something valuable for agents is the fact they desire it.<sup>2</sup> On such a theory, being desired and being good are distinct properties: the former is natural and non-normative while the latter is genuinely normative and reason-implicating. The fact that something is desired is supposed to confer to it a distinct normative property: precisely in virtue of being desired, it becomes *desirable*, and the agent acquires normative reasons to favour it in some way.

<sup>1</sup> Zarka (1996).

<sup>2</sup> Gauthier (1979); Hampton (1986: 17, 27–42, 53); Ripstein (1987: 124); Meyers (2013: 273).

If Hobbes subscribed to subjectivist, desire-based theories of reasons and value, then we might have a plausible explanation for how ethics, despite being irreducibly normative, could primarily consist in the study of human psychology. To be sure, ethics would include foundational irreducibly normative principles, according to which the content of one's desires is the source of reasons and value. But it would otherwise amount to the study of human psychology, because which *particular* reasons agents have, and which particular things have value for them, would be determined by the "motions" of their mind. Moreover, assuming the normative principles of subjectivism could be known *a priori*, the subjectivist interpretation would seem able to explain how humans could come to know what reasons they have and which particular things have value for them. For Hobbes, passions are not merely conative but also cognitive evaluative states that represent their intentional objects as good or evil. If desiring something makes it normatively desirable for agents or gives them a reason to favour it, then the intentional content of desire is a perfect medium by which to know which objects furnish reasons or have value.

Yet in fact Hobbes was not a subjectivist about either reasons or value. The next two chapters vindicate this thesis.

## *Subjectivism, Instrumentalism, and Prudentialism about Reasons*

When Aquinas fused Aristotle with Christianity to cast natural-law theory into its classic mould, he embedded it firmly within the philosopher's conception of science. Aristotle had asserted that because the first premises or principles (*arkhē*) of science (*epistēmē*) are indemonstrable, they cannot themselves be the object of inferential, scientific knowledge: their truth must be cognized non-inferentially by the intellect (*nous*).<sup>3</sup> Aquinas echoed this point: the indemonstrable foundations of science, i.e., the very "first universal principles [*principia*] . . . from which reasoning begins, both in speculative and in practical matters," must become "known to us by the natural light of the intellect."<sup>4</sup> Since scientific knowledge is a matter of inferential reasoning, each science must take its own starting premises for granted:<sup>5</sup> the "principles of any science are either in themselves self-evident [*nota per se*], or reducible to the conclusions of a higher science"<sup>6</sup> – where a proposition is self-evident "whose predicate is contained in the notion [*ratione*] of the subject."<sup>7</sup> There are therefore two kinds of truth: "what is known in itself [*per se notum*] . . . as a *principium*, and at once understood by the intellect," and truths known "not at once, but through rational inquiry, and as the terminating point [*termini*] of reasoning."<sup>8</sup> As much as Hobbes repudiated scholastic teleology, he retained key elements of its foundationalist picture of science: he too agreed that scientific reasoning must begin with self-evident first principles.

Aristotle had claimed in *Metaphysics* 4.9 that in speculative reasoning, as Aquinas put it, the "first indemonstrable principle" is non-contradiction, i.e., that "*the same thing cannot be affirmed and denied at the same time.*" Aquinas endorsed this foundational principle, but also drew a parallel with practical reasoning: "the precepts of natural law are to practical reasoning what the first principles of demonstration are to speculative reasoning,

<sup>3</sup> NE 7.6–7.    <sup>4</sup> ST I-II.62.3.    <sup>5</sup> ST I.1.8.    <sup>6</sup> ST I.1.2.    <sup>7</sup> ST I-II.94.2; cf. I.2.1; I-II.66.5.

<sup>8</sup> ST I-II.57.2.

insofar as both are *principia per se nota*.” In particular, because “every agent acts for the sake of an end, which has the nature of a good [*rationem boni*],” the first principle of practical reasoning is that “the good is what all things desire.” This does not yet quite seem a normative principle: it is one thing for Aquinas to tell us we are inclined to seek the good, and even to specify what our ultimate good or felicity consists in; it is another to say we ought or have a normative reason to pursue it. The question is whether felicity or the good is a genuinely normative concept; what is required is something linking Aquinas’s theory of the *good* to his theory of *reasons*. Precisely this link is furnished by the “first precept of law” which serves as the foundation of all natural-law precepts: that “*good is to be done and pursued*.” This foundationally normative precept, and our capacity to know its truth, derives from the world’s teleological nature for Aquinas: because nature is a rational, normative order in which each being has a natural *telos* or end, “all those things to which man has a natural inclination, are naturally apprehended by reason as being good, and consequently as objects of pursuit.”<sup>9</sup> Thus when Aquinas stated that “in actions, which are the object of practical reasoning, the first principle is the ultimate end . . . [or] felicity,” he took this to be not merely a descriptive, but genuinely normative point. And since he assumed that the “*felicitem communem*” or common good is constitutive of each individual’s felicity, natural law is foundationally “ordained to the common good.”<sup>10</sup>

Humans are naturally inclined, according to Aquinas, to self-preservation; to preserving the species through sexual union and childrearing; and, because of our rational nature, to knowledge of God and to social life. It is our rational – and hence theological and sociable – nature that helps one to discern the common good as constitutive of one’s own. Hence even in reflecting on our inclination to self-preservation, one sees that natural laws cover the means to “preserving human life” as such, rather than just one’s own life.<sup>11</sup> From the inclinations of our rational nature we can thus know two self-evident general precepts of natural law, which are corollaries to pursuing the good: namely, to love God and our fellows. Natural law contains three grades of moral precept: first, self-evident precepts such as these two (which serve as foundational principles for the Decalogue’s more determinate moral precepts); second, derivative precepts everyone can grasp but that require promulgation to avoid errors (which are revealed in the Decalogue); and third, precepts manifest to the wise but not to the many.<sup>12</sup>

<sup>9</sup> ST I-II.94.2.   <sup>10</sup> ST I-II.90.2.   <sup>11</sup> ST I-II.94.2.

<sup>12</sup> ST I-II.100.3; 100.11. Cf. Suárez, DLDL I.3;18; 7.4; II.7.5–7; 8.2–4; 15.2.

Hobbes transformed this Thomist theory of natural law in two critical ways. Despite retaining Aquinas's foundationalism, he dropped the teleological cosmology, and repudiated the assumption of sociability and hence the foundationally common-good orientation – thereby grounding natural law purely in the individual's own good and preservation. Both transformations have misled Hobbes's modern interpreters. His desertion of a teleological cosmology has led many to see him as a subjectivist about reasons (or, at the metaethical level, as an ethical-naturalist reductionist). And his focus on self-preservation has led many to think Hobbes abandoned the notion of an ultimate good or felicity – if not entirely, then at least as the normative foundation of natural law and thus of his theory of reasons. My aim here is to show that Hobbes did not advance a subjectivist or even instrumentalist theory, but rather a prudentialist theory of natural law. And the self-evident first principles of the science of ethics – which consist in definitions of the most basic normative terms, such as 'good' and 'natural law' – ground the laws of nature firmly in one's own felicity.

### 3.1 Conative Subjectivism about Reasons? The Problem of Instrumental Transmission

If one has a normative reason to do anything, it is to desire and secure one's own preservation. That, at any rate, is the minimal normative claim frequently attributed to Hobbes – as the foundational principle of natural law.<sup>13</sup> "A LAW OF NATURE," Hobbes declared in *Leviathan*, "is a Precept, or generall Rule, found out by Reason, by which a man is forbidden to do, that which is destructive of his life, or taketh away the means of preserving the same."<sup>14</sup> That one ought to desire self-preservation is not itself "found out" by reasoning, but is the principle or starting point of practical reasoning embedded in the definition of natural law; through reasoning we discover general rules about which actions are the relevant means to self-preservation. And since Hobbes's first law of nature prescribes that each (be disposed to) seek peace, and the secondary laws prescribe the relevant means to peace, I take a dual touchstone for any adequate interpretation of Hobbes to be its ability to explain the universal normativity<sup>15</sup> not just of (1) self-preservation, but also of (2) peace and its relevant means.

Hobbes's definition notably abandons any foundational orientation to the common good. Indeed, contrary to my thesis, it appears to sidestep

<sup>13</sup> Tuck (1989, 1996). <sup>14</sup> L 14.3: 198.

<sup>15</sup> Natural law addresses "each and every man [*unumquemque*]." DCv 3.1.

even one's own good or felicity in favour of an exclusive focus on self-preservation.<sup>16</sup> This widely alleged contrast between self-preservation and felicity is the linchpin of the so-called standard interpretation of Hobbes. On this interpretation, Hobbes viewed the desire for self-preservation as the counterpart to the fear of death and hence equivalent to a desire for ongoing *survival*, and concluded that humans have a decisive reason to desire and secure survival on the basis of two further premises. First, the aversion to death and desire for survival are overriding human passions; and second, the fact we actually desire something furnishes a reason to desire and secure it – a reason whose strength is proportional to the desire's intensity. To be sure, Hobbes did not view survival as a *telos* inscribed by God into the fabric of the universe: there are no ontologically objective purposes in nature on his view.<sup>17</sup> Rather, our own *desire* is supposed to confer normativity to survival. As Hobbes put it in *De Cive*,

every man is desirous of what is good for him, and shuns what is evil, but chiefly the chiefest of naturall evils, which is Death; and this he doth, by a certain necessity of nature, no lesse then that whereby a Stone moves downward. It is therefore neither absurd, nor reprehensible; neither against right reason for a man to expend all his efforts to preserve and defend his Body, and the Members thereof from death and sorrowes.<sup>18</sup>

Since we are averse to death above all, we ought to desire and secure our own survival. This, in turn, is supposed to be the principle grounding the universal normativity of natural law. The standard interpretation thus attributes to Hobbes, at the empirical level, a psychology according to which there is a universally overriding passion and, at the normative level, conative subjectivism about affective and practical reasons.<sup>19</sup>

Both premises are mistaken. The problem with the empirical premise is that Hobbes denied the fear of death is always decisive. It is frequently overridden by the desire for revenge and glory (when subjected to “hatred, or contempt . . . most men choose rather to hazard their life, than not to be revenged”<sup>20</sup>); by fear of imaginary things, such as ghosts or eternal damnation;<sup>21</sup> and by desires the satisfaction of which is closer at hand.<sup>22</sup> To be sure, according to Hobbes the fear of death is a natural and especially

<sup>16</sup> Schneewind (1998); Garrett (2013). <sup>17</sup> DC 10.7.

<sup>18</sup> DCv 1.7. See EL 14.6; L 13.3: 190; 15.17: 232.

<sup>19</sup> The standard interpretation is so-named and defended by Murphy (2000). See Hampton (1986); Lloyd (1992: chapter 1).

<sup>20</sup> L 15.20: 234. Cf. DCv 3.12. This observation is not, as Murphy suggests, peripheral: it is central to Hobbes's account of war (Abizadeh 2011).

<sup>21</sup> L 29.15 : 510–512; 42.70: 854. Lloyd (1992). <sup>22</sup> DCv 3.31–32.

strong aversion, but just as a stone's natural impulsion downwards can be thwarted by contrary forces, so too can the aversion to death be overridden by other passions. The standard view consequently has no explanation for why, as Hobbes thought, natural law binds even those vengeful glory-seekers, ghost- and damnation-fearers, and myopes for whom self-destructive passions are stronger than all other passions.

To see the problem with attributing a conative-subjectivist theory of affective and practical reasons to Hobbes, begin by considering the theory in its purest form. *Pure conative subjectivism* grounds affective and practical reasons in facts about the content of agents' present desires. Consider the following twin precepts:

The *Subjective Precept for Desire* (SP<sub>D</sub>) prescribes that: one presently desire some event or state of affairs if (and in virtue of the fact that) one presently desires that event or state of affairs;

The *Subjective Precept for Action* (SP<sub>A</sub>) prescribes that: one presently undertake some action if (and in virtue of the fact that) one presently desires to undertake the action;

where the strength of each prescription is proportional to the strength of one's present desire.

Pure conative subjectivism asserts that these twin precepts are *a priori* normative principles, and that no other precept is normative unless derivable from them. We can put this by saying that SP are the only foundationally normative desire- and action-precepts. If pure conative subjectivism is true, then one presently has a reason to desire something if and only if (and in virtue of the fact that) one presently desires it, and one presently has a reason for action if and only if (and in virtue of the fact that) one presently desires to undertake it. So the fact you presently desire to visit your grandmother would give you an affective reason to desire your visiting with her, and a practical reason to visit her.

The problem is that pure subjectivism rules out the normativity of the prescription to agents that, if they have a reason to desire some event or state of affairs, or a reason to undertake some action, they desire or take the relevant means to it. Consider the following Instrumental-Transmission Precepts (ITP):

The *Instrumental-Transmission Precept for Desire* (ITP<sub>D</sub>) prescribes that: one presently desire some event or state of affairs if (and in virtue of the fact that) it is a relevant means to some event or state affairs one presently has a reason to desire or to some action one presently has a reason to undertake;

The *Instrumental-Transmission Precept for Action* (ITP<sub>A</sub>) prescribes that: one presently undertake some action if (and in virtue of the fact that) it is

a relevant means to some event or state of affairs one presently has a reason to desire or to some action one presently has a reason to undertake;  
 where the strength of each prescription is proportional to the strength of the supposed reason and the relevance of the means;  
 and where means include causal conditions, criterial connections, facilitating conditions, and constitutive conditions.<sup>23</sup>

If ITP are genuinely normative, then agents have a reason to do what they prescribe: to desire and to take the relevant means to their normative ends. Such precepts would transfer reasons from normative ends to relevant means.<sup>24</sup> So you have a practical reason to *visit* your grandmother if you have an affective reason to *desire* that you visit her; and you have a practical reason to leave your house if you have a practical reason to visit your grandmother.

That pure conative subjectivism rules out the normativity of Instrumental Transmission becomes apparent as soon as we confront agents who desire some end, but do not desire to take the relevant means to it (and do not desire to conform to ITP). Precisely because they do not desire to take the means (nor to conform to ITP), on the pure conative-subjectivist theory they have *no* normative reason to do so – even though the end in question is normative for them. You may have a reason to visit with your grandmother (in virtue of your desire to do so), but you would have no reason to leave your house unless you actually desired to leave your house.

The reason why pure conative subjectivism is implausible as an interpretation of Hobbes is apparent. Even if pure conative subjectivism could account for the normativity of self-preservation, it would fail to establish the normativity of the laws of nature, because these strictly prescribe (being disposed to) taking the relevant *means* of self-preservation. Granting that each person has a reason to preserve his own life – which Hobbes called “the reason of his preservation”<sup>25</sup> – each also has a reason to adhere to the first law of nature – “*to seek Peace*” – precisely because it is the relevant “means of preserving” each person’s own life; and each has a reason to adhere to the subsequent laws of nature because they, in turn, are the relevant “means of peaceable . . . living.”<sup>26</sup> But on the pure subjectivist

<sup>23</sup> See Hubin (1999: 32); Raz (2005). For finer-tuned versions of these precepts, see Abizadeh (in progress).

<sup>24</sup> If ITP are normative, they are synthetic normative principles, and so differ from the “Transmission principle” discussed in Skorupski (2010: 103–104), which is analytic. Cf. Broome (2013: 126); Kiesewetter (2015). ITP are distinct from rational requirements concerning relations between mental states (Broome 2005).

<sup>25</sup> L 15.5: 224. <sup>26</sup> L 14.4: 200; 14.3: 198; 15.40: 242; cf. 15.36: 240.

theory, even if one has an overriding aversion to death (as the standard interpretation claims everyone has), but fails to desire peace, fails to desire to seek peace, or fails to desire to avoid cruelty or pride – as prescribed by the seventh and ninth laws of nature<sup>27</sup> – then one would have no reason to do so.

Hobbes, by contrast, explicitly endorsed the normativity of Instrumental Transmission: the normativity of the secondary laws of nature, derived from the first, is underwritten by Instrumental Transmission. His endorsement occurs in *De Cive* in the course of deriving the secondary law of nature declaring

*That safety must be assured to the Mediators of Peace.* For the reason which enjoins the end, enjoins also the necessary means [*Ratio enim, quae iubet finem, iubet etiam media necessaria*]. But the first dictate of Reason is *Peace*; All the rest are means to obtain it, and without which *Peace* cannot be had. But neither can *Peace* be had without mediation, nor mediation without safety; it is therefore a dictate of Reason, that is a *Law of nature*, That we must give all security to Mediators of *Peace*.<sup>28</sup>

(Hobbes made the same point about the Instrumental Transmission of goodness: if “Peace is Good,” then, as he put it *Leviathan*, “therefore also the way, or means of Peace” are also good, such that all “the Laws of Nature, are good,” and that “wherein consisted their Goodnesse” is the fact they are the relevant “meanes” to peace.<sup>29</sup>)

If the normativity of ITP requires abandoning pure conative subjectivism, it is because these precepts prescribe desires and actions solely on the basis of mind-independent facts *about the world* – about what events or states of affairs serve as means to other events or states of affairs – rather than on the basis of facts about agents’ desires. They are conatively *objective* precepts: their prescriptions are not relativized to (the content of) agents’ desires.

Partisans of the standard interpretation might seek to salvage their reading by imputing to Hobbes a modified conative subjectivism grounding normative reasons not in agents’ *actual* present desires, but in the desires they hypothetically *would* have if they had reasoned (or deliberated) rightly. The empirical premise of this revised view is not that agents universally *do* have an overriding desire for self-preservation, but that they would do so upon reasoning (or deliberating) rightly. There would consequently be no need to claim the normativity of Instrumental Transmission: one could assert instead that for Hobbes it is a *psychological fact* that when agents reason

<sup>27</sup> L 15.19–21: 232–234.

<sup>28</sup> DCv 3.19. Cf. L 15.29: 236.

<sup>29</sup> L 15.40: 242. Cf. DCv 3.31.

rightly, their *desires* (rather than normative reasons) are transferred to the relevant means. Hobbes's apparent endorsement of Instrumental Transmission, on this revised view, would amount to an empirical claim about the results of *reasoning*, rather than a normative claim about what *reasons* agents have: as Hobbes also put it, "every man by reasoning [*rationcinando*] seeks out the meanes to the end which he propounds to himselfe."<sup>30</sup> At the normative level, the revised view embraces a *reasoning-based* (or a healthy-deliberation-based<sup>31</sup>) theory of affective and practical reasons, which claims that the following precepts, rather than SP, are the only foundationally normative principles:

The *Reasoning-Based Precept for Desire* (RP<sub>D</sub>) prescribes that: one presently desire some event or state of affairs if (and in virtue of the fact that) one would presently desire that event or state of affairs if one had reasoned rightly (and/or deliberated healthily);

The *Reasoning-Based Precept for Action* (RP<sub>A</sub>) prescribes that: one presently undertake some action if (and in virtue of the fact that) one would presently desire to undertake the action if one had reasoned rightly (and/or deliberated healthily);

where the strength of each prescription is proportional to one's hypothetical present desire.

On the reasoning-based theory, no other precept is normative unless derivable from these two.

Let us initially grant the first, empirical part of this revised interpretation (that agents would overwhelmingly desire survival upon reasoning rightly). The question is what justifies, from a conative-subjectivist perspective, the claim that agents have a reason to favour something if (and only if) it is the object of their hypothetical, reasoning-based desires. There are two possible answers: either right reasoning *confers* normativity to the object of the hypothetical desires, or these objects are independently normative, and the hypothetical desires are merely a reliable heuristic tool for merely *discovering* or tracking the reasons one has.

The first path is pursued by Mark Murphy, who seeks to salvage the standard interpretation in the face the fact that, according to Hobbes, the fear of death is frequently overwhelmed by potentially self-destructive passions. Murphy argues that the greater strength of some self-destructive desires may arise from the failure to represent in the imagination future states of affairs (such as one's death) in a sufficiently vivid way and, more generally, that the failure decisively to desire survival invariably

<sup>30</sup> DCv 14.16.    <sup>31</sup> Hampton (1986: 38–40), drawing on Parfit (1984: 118).

results from deficiencies in deliberation. He further asserts that the subjectivist may discount the normativity of desires arising from “cognitive deficiency” or “unhealthy” deliberation: drawing on Jean Hampton’s early work on Hobbes, he argues that only desires resulting from “healthy” deliberation furnish normative reasons.<sup>32</sup>

But Murphy’s argument begs the question of why representing more distant events less vividly counts as a “deficiency” and – as Hampton recognized in later work – from where the norm of healthy deliberation arises.<sup>33</sup> The problem is that while this first path is compatible with RP being the only foundationally normative principles, it entails that they are, like ITP, conatively *objective* – according to which the standards of right reasoning furnish normative reasons. This not only departs from conative subjectivism, it is also implausible as an interpretation of Hobbes: Hobbes characterized right reasoning not as a process that *confers* reasons, but as a process by which agents would *discover* them. The laws of nature are “found out by Reason.”<sup>34</sup>

The second option, according to which by reasoning one discovers reasons, concedes this, but precisely because it claims that hypothetical, reasoning-based desires *track* our reasons, it is *not* compatible with claiming RP are normative. It implies reasons are normative not *in virtue of the fact that* something would hypothetically be desired, but in virtue of other facts. The second option therefore presupposes the truth of some other set of normative principles that explain why some facts about the world (not agents’ desires) furnish normative reasons. Hence while the second option is compatible with the normativity of SP, it also presupposes, like the first option, a set of objective, rather than solely conative-subjective foundationally normative principles. And if SP are normative, the most obvious principles to accompany them is ITP. No wonder most commentators attribute to Hobbes an *instrumentalist* theory of affective and/or practical reasons.

### 3.2 Instrumentalism about Reasons? The Problem of Time

*Instrumentalism* about affective and practical reasons asserts the joint but exclusive foundational normativity of SP and ITP.<sup>35</sup> This combination entails the normativity of the following:

<sup>32</sup> Murphy (2000: 54–57). <sup>33</sup> Hampton (1992). Cf. Parfit (2011: vol. 1, 93–95). <sup>34</sup> L 14.3: 198.

<sup>35</sup> Hubin (1999). Schroeder’s (2007) “hypotheticalism” is no exception. Hypotheticalism provides an interpretation of why, according to IP, one has a reason for action *in virtue of the fact* that it is a relevant means: the fact that one desires something does not *constitute* the reason, but is

The *Instrumentalist Precept for Desire* (IP<sub>D</sub>) prescribes that: one presently desire some event or state of affairs if and only if (and in virtue of the fact that) it is a relevant means to an event or state affairs one presently desires;

The *Instrumentalist Precept for Action* (IP<sub>A</sub>) prescribes that: one presently undertake some action if and only if (and in virtue of the fact that) it is a relevant means to some event or state of affairs one presently desires;

where the strength of each prescription is proportional to the strength of one's present desire and the relevance of the means.

Instrumentalism might seem to provide a promising explanation of the normative status of self-preservation and natural law in Hobbes. The instrumentalist interpretation not only can acknowledge the normativity of ITP, it raises the possibility of asserting that death is a universal evil everyone ought to avoid, and ongoing survival a universal good each ought to promote, without having to claim that everyone has an overriding aversion to death. One might instead claim that being alive is the relevant means for promoting *whatever else* an agent may desire, and then derive the normativity of self-preservation on this empirical basis in conjunction with the normativity of IP. David Boonin, for example, argues that for Hobbes "death is the elimination of all future desire-satisfaction, and is thus the greatest evil for each individual."<sup>36</sup> Rational agents therefore universally have a decisive normative reason to desire and secure their own survival. This is in turn supposed to explain why the laws of nature bind even a person who happens to desire other things more than self-preservation or peace.

The instrumentalist interpretation faces two objections, however: one against its empirical premise (that survival is the relevant means to the object of all other desires), another against its normative premise (IP). The problem with the empirical premise is that death does not always prevent the satisfaction of one's present passions; nor does its prospect inhibit all anticipatory pleasure. It is simply not true that survival is always a relevant means to the object of any passion – and Hobbes knew this.

First, survival is not always a relevant means to fulfilling *post-mortem* passions, i.e., desires for and aversions to what happens after one's death. Their fulfilment *presupposes* one's death. Second, it is not always a relevant means to fulfilling *self-destructive* passions, i.e., desires and aversions such that either the attempt to fulfil them, or their satisfaction, involves or would likely lead to one's death. Beyond individuals who presently desire

a background condition explaining why some *other* fact constitutes a reason. Either way, IP incorporates one subjective and one objective element (Abizadeh in progress).

<sup>36</sup> Boonin-Vail (1994: 55–56). Cf. Strauss (1963: 15–18).

to die, consider those for whom the aversion to death is weaker than their self-destructive desire to take unlawful (capitally sanctioned) revenge,<sup>37</sup> to martyr their life for a religious cause,<sup>38</sup> or to avoid the infamy of killing their own parent (whom they have been commanded to kill, on pain of death, by their sovereign).<sup>39</sup> These people's death need not preclude the fulfilment of these stronger, self-destructive desires.<sup>40</sup> The instrumentalist interpretation is incapable of explaining how natural laws could be binding for agents with (strong) post-mortem and self-destructive passions.<sup>41</sup>

To see the problem with the instrumentalist interpretation's normative premise, consider what could justify combining its two distinct elements – one subjective (SP), the other objective (ITP) – into a single theory of reasons, rather than abandoning SP altogether. It might be thought this challenge could be met by appealing to a theory of well-being. The proposal here is to justify the combination of SP and ITP into a single theory of normative *reasons* insofar as both elements are commonly related to an agent's *good*. On this value-based proposal, the truth (articulated by SP) that one has a normative reason to undertake some action if one desires to do so would be explained by reference to the fact that (one can reasonably know or expect that) fulfilling the desire to so act will promote one's well-being. The truth (articulated by ITP) that one has a normative reason to take the relevant means to one's normative end would be explained in the same way: because (one can reasonably know or expect that) taking the relevant means will promote one's well-being. The textual basis for such a proposal would presumably lie in Hobbes's invocation of "that Reason, which dictateth to every man his own good."<sup>42</sup>

The proposal might be framed in relation to either of two theories of the good typically attributed to Hobbes, namely, the desire-fulfilment and hedonist theories. On the desire-fulfilment theory, the good of agents consists in the objective satisfaction of their desires and objective non-frustration of their aversions. On the hedonist theory, the good of agents consists in pleasure and absence of pain: desire fulfilment may contribute to well-being, but only instrumentally, insofar as being aware of or anticipating it yields pleasure. For present purposes, it does not matter which interpretation is correct: either way, in typical cases one can reasonably

<sup>37</sup> DCv 3.12; L 15.20: 234.   <sup>38</sup> DCv 18.13; L 42.12–14: 786–788.   <sup>39</sup> DCv 6.13.

<sup>40</sup> Although McNeilly (1968: 181) acknowledges that survival is not a relevant means for fulfilling every desire, he falsely suggests peace is (191). Consider a desire for war or post-mortem martial fame. For a different criticism, see Lloyd (2009: 208–209).

<sup>41</sup> Murphy (2000: 49) rejects the "normative" interpretation of self-preservation on similar grounds.

<sup>42</sup> L 15.4: 222.

know or foresee that acting to fulfil one's desires will contribute to one's well-being – whether intrinsically or instrumentally.

But the value-based proposal cannot support instrumentalism. The core problem is that while the future fulfilment of one's desire may contribute to one's *future* well-being, it cannot itself contribute to one's *present* well-being. That, at any rate, is what follows from a *concurrentist* theory,<sup>43</sup> and we must attribute concurrentism to Hobbes whatever his theory of well-being.<sup>44</sup> On a concurrentist theory, one's well-being at any point in time is intrinsically enhanced (diminished) by, and only by, the fulfilment (frustration) *at that time* of the desires (aversions) one has *at that time* or, in its hedonist variety, only by one's pleasures (pains) *at that time*. Hobbes's concurrentism is illustrated by his treatment of the post-mortem "Desire for Fame after death." Hobbes assumed that fulfilling a post-mortem desire cannot contribute to one's present well-being because, by definition, it cannot be presently fulfilled, and no present pleasure can be retroactively caused by the desire's future fulfilment. To be sure, one's *desire* for post-mortem fame can contribute to one's present good insofar as one takes "a present delight therein, from the foresight of" the future fame, that is, insofar as one experiences an anticipatory "pleasure in the imagination."<sup>45</sup> On the desire-fulfilment theory, this pleasure contributes to one's present good insofar as one presently desires pleasure; on the hedonist theory, the pleasure enhances one's present good intrinsically. But the *objective satisfaction* of one's present desire for post-mortem fame – which can only obtain in the future – itself contributes neither to one's present good (since it is not yet satisfied) nor to one's future good (since for Hobbes death terminates earthly felicity).<sup>46</sup>

This confronts instrumentalism with a basic challenge: Why should the fact that some present action is a relevant means to the *future* satisfaction of one's present desire give one a *present* reason to undertake it, given that it cannot enhance one's present well-being? The only answer available to the value-based proposal is that one has a reason to care about and promote one's future well-being – regardless of whether one presently does care or not. Yet *this* reason does not intrinsically arise from the normativity of IP. IP merely prescribes undertaking actions that are the relevant means to events or states one presently does desire.

<sup>43</sup> Heathwood (2005); Sarch (2013); Lin (2018).    <sup>44</sup> Cf. Meyers (2013).

<sup>45</sup> L 11.6: 152. See DH 12.8.    <sup>46</sup> On such cases, see Nagel (1970).

Consider an out-of-shape individual who presently desires that he were physically fit and who reasonably and correctly judges that at all points in the future he will desire to be physically fit, that being fit would enhance his well-being, and that some present actions, such as exercising, are relevant means for fulfilling his ongoing desire. Imagine, however, that his awareness of all these truths nevertheless fails presently to give rise to any desire to take these relevant means: he does not desire to exercise. This may happen, indeed according to Hobbes it frequently happens, because even if he were to take the relevant means, his ongoing desire would not be satisfied, and his good enhanced, until the future; and he may at present just not care enough about his future. Perhaps he is a Cyrenaic, or suffers from intemperance. Precisely because “*things present* are obvious to the sense, [and] *things future* to our Reason only,” humans frequently succumb to an “irrational appetite, whereby they greedily prefer the present good . . . before the future.”<sup>47</sup> Present actions geared to future results are paradigmatic of instrumental action: causal means temporally precede the end. That one nevertheless has a reason *presently* to take the relevant means to fulfil, in the future, some ongoing desire presupposes that one has a reason to care about or promote one’s future well-being.

### 3.3 Prudentialism about Reasons

The problem of time thrusts into view the normativity of precepts of prudence,<sup>48</sup> which prescribe that agents care about and act to secure their ongoing, overall well-being or felicity. The problem of time reveals that the value-based proposal, far from supporting an instrumentalist union of ITP and SP, supports instead combining Instrumental Transmission with the following:

The *Pure Prudential Precept for Desire* (PPP<sub>D</sub>) prescribes that: one presently desire some event or state of affairs if (and in virtue of the fact that) it is a constituent of one’s ongoing felicity;

The *Pure Prudential Precept for Action* (PPP<sub>A</sub>) prescribes that: one presently undertake some action if (and in virtue of the fact that) the action is a constituent of one’s ongoing felicity;

where the strength of each prescription is proportional to the amount of good at stake and to some future-utility-discount rate  $d$  ( $1 > d$ ).<sup>49</sup>

If ITP and PPP are both normative, one can also derive the normativity of:

<sup>47</sup> DCv 3.31–32; cf. DH 12.1.   <sup>48</sup> See Nagel (1970: 36); Korsgaard (2008).

<sup>49</sup> On a future-utility-discount rate, see “pure” discounting in Broome (1994).

The *Prudential Precept for Desire* (PP<sub>D</sub>) prescribes that: one presently desire some event or state of affairs if (and in virtue of the fact that) it is a relevant means to an event or state affairs that is good for one;

The *Prudential Precept for Action* (PP<sub>A</sub>) prescribes that: one presently undertake some action if (and in virtue of the fact that) it is a relevant means to some event or state of affairs that is good for one;

where the strength of each prescription is proportional to the amount of good at stake, the relevance of the means, and some future-utility-discount rate  $d$  ( $1 > d$ ).

*Ethical egoism* asserts that PP (or PPP and ITP jointly) are the sole foundationally normative precepts. According to egoism, one presently has a reason to desire or do something if and only if, and in virtue of the fact that, it is a relevant means to one's own overall good. This is the sense in which the classical, eudaimonistic theory of reasons is an egoist theory. *Prudentialism*, by contrast, asserts that PP (or PPP and ITP) are foundationally normative (and that the discount rate  $d$  is sufficiently low), but potentially leaves the door open to other foundationally normative precepts.

I take it that whatever else Hobbes's ethics is, it is prudentialist: Hobbes's invocation of "that Reason, which dictateth to every man his own good" is an endorsement of the normativity of PP. Yet Hobbes was not an ethical egoist: not *all* normative reasons are grounded in one's own good. As I shall argue in [Part III](#), Hobbes acknowledged reasons of the right not derivable from reasons of the good. But Hobbes did take all reasons *of the good* to be *prudential* reasons centred on the agent's own good. Indeed, as the passage concerning "irrationall appetite" cited earlier indicates, Hobbes embraced *strong prudentialism*, which sets the discount rate  $d$  to nil: the amount of good that determines the relative *weight* of one's reasons to favour some means is the net good at stake in the entire chain of consequences; the weight of each link is independent of its place in the future temporal chain.

My thesis is that the normativity of self-preservation and natural law in Hobbes presupposes, and cannot be explained except by appeal to, his prudentialism. The normativity of self-preservation and the laws of nature, which strictly prescribe that agents each (be disposed to) preserve themselves and to take the relevant means to self-preservation, presupposes that agents presently have a reason to care for their ongoing, overall good – their felicity – and not just their present good. If for Aquinas the first normative principle which grounds natural law is *good is to be done and pursued*, for Hobbes it is *desire and pursue your own good*. It is true, of course, that

humans are wired, according to Hobbes, to pursue what *appears* good to them; but as we shall see, what appears good – i.e., the object of present desire – frequently fails to be truly conducive to one’s felicity. This is the sense in which Hobbesian natural laws are accurately described as precepts or “maxims” of prudence.

Several objections to my thesis are worth considering: that (1) my use of ‘prudence’ in this sense is alien to Hobbes; (2) the laws of nature are supposed to be discovered by reasoning and hence are a part of science, which Hobbes sharply contrasted to prudence; (3) reasoning for Hobbes is formal and so is not of ends; and (4) as a purported explanation of the universal normativity of self-preservation and the laws of nature, prudentialism is vulnerable to an empirical objection akin to the one raised earlier against instrumentalism, namely that survival is not the relevant means for satisfying all passions. I take up the first three objections here, turning to the fourth in this chapter’s final section (where I argue that for Hobbes self-preservation is not equivalent to survival). A fifth objection, that the laws of nature furnish not merely prudential reasons of the good but also reasons of the right, I take up in [Part III](#).

The first objection is that the notion of prudence at stake in prudentialism, which concerns *affective* reasons for desire and *practical* reasons for action, is alien to Hobbes. Hobbes characterized prudence as consisting in cautious conjecture, on the basis of past experience, about the likely consequences of present events or actions; he in turn portrayed the excellence or virtue of prudence as comprising greater accuracy in conjecturing, thanks to much experience. Thus in *Elements* he wrote that prudence is “nothing else but Conjecture from Experience . . . warily” about what consequences are “likely to come to passe, or to have passed allready” in the “Past & Future,” and “that they shall conjecture best that have most experience.”<sup>50</sup> One might therefore conclude that for Hobbes prudence is concerned not with affective or practical reasons, but with *epistemic* reasons for belief, namely beliefs about what is likely to occur in the future; it is not, as prudentialism is, centred on an agent’s felicity.

The objection has no teeth. The alleged fact that Hobbes used ‘prudence’ solely in an epistemic sense would not, even if true, preclude him from endorsing the thesis that agents’ felicity is normative for them. (Nor would it preclude us from characterizing such a thesis as “prudentialist” in an affective/practical sense of the term.) In any case the alleged fact is not true. Hobbes used the term ‘prudence’ in two senses: both (a) the epistemic

<sup>50</sup> EL 4.10–11.

sense and (b) the affective/practical sense centred on an agent's felicity. Two years after completing *Elements*, for example, Hobbes in *De Cive* characterized failures to take the relevant means to one's own ongoing good – illustrated by those hapless fools who build their house “upon the Sand” – as failures “contrary to right reason” amounting to “sins of imprudence [*peccata imprudentiæ*].”<sup>51</sup> Shortly thereafter, in *Anti-White*, he linked the two (epistemic and affective/practical) senses of prudence. He defined prudence as foresight, based on experience of consequences, of the way to felicity, and characterized excellence in prudence as accuracy in determining the way of felicity:

prudence consists in foresight of the path to felicity . . . Moreover, foresight of the future consists in the memory or experience of past consequences, of which similar ones are expected in the future, thus the most prudent are those who have the most experience, and the most experienced are those who have acquired many observations about the consequences of things good and bad.<sup>52</sup>

Hobbes linked the two senses of prudence again in *Leviathan*. He there defined prudence in epistemic terms as “foresight of the things to come” and as “a *Praesumption* of the *Future*, contracted from the *Experience* of time *Past*.” But in between these statements he folded in a claim that uses the term in both its (a) epistemic and (b) affective/practical senses: “it is not Prudence that distinguisheth man from beast. There be beasts, that at a year old [a] observe more, and [b] pursue that which is for their good, more prudently, than a child can do at ten.”<sup>53</sup> Fully prudent agents not only *foresee* accurately, but also *pursue* their own long-term good; they are prudent in both the epistemic and affective/practical senses. To say that the laws of nature are (ultimately grounded in) precepts of prudence is to use the term in its affective/practical, not epistemic, sense.

This distinction, between the two senses of prudence, also dissolves the second objection. It is true that Hobbes sharply distinguished between prudence and empirical observation, on the one hand, and science and reasoning, on the other. Prudence, which “Brute Beasts” can have as well, depends on “knowledge of Fact” obtained by empirical observation: this type of knowledge “is originally Sense; and ever after, Memory.” By contrast, scientific knowledge or the “knowledge of Consequence” is restricted to humans and is obtained by commencing with definitions treated as principles and then reasoning syllogistically in verbal discourse

<sup>51</sup> DCv 14.16; 14.19.    <sup>52</sup> AW 38.9. See AW 38.4.    <sup>53</sup> L 3.7–10: 44.

to derive conclusions.<sup>54</sup> Natural laws are the object of scientific rather than factual or prudential knowledge: they are “dictates of Reason,” and the “true Doctrine of the Lawes of Nature” composes the “Science” of “true Morall Philosophie.”<sup>55</sup> It might therefore be objected that, because the laws of nature are the objects of scientific rather than prudential knowledge, they cannot constitute precepts of prudence.<sup>56</sup> But the objection depends on an equivocation between the two senses of prudence. The objection’s premise is correct: the laws of nature are known scientifically via reasoning, not via prudence in the *epistemic* sense of the term; they are not discovered via merely empirical observation and foresight, but “found out by Reason.”<sup>57</sup> The objection’s conclusion does not follow, however, because in saying that natural laws are prudential precepts we are using the term in the affective/practical sense: we are not indicating how such laws are known; we are saying they are grounded in the prescription to care for and secure one’s felicity. Hobbes did not think that acquiring prudence in the affective/practical sense relies solely on epistemic prudence. That it also depends on scientific reasoning is why a “Science of what is *Good*”<sup>58</sup> is necessary. So even if the laws of nature, insofar as they are grounded in the prescription to care for and take the relevant means to felicity, are prudential in the affective/practical sense, they can nevertheless constitute a true science of what to (be disposed to) do: the means they prescribe, such as peace and upholding covenants, can be scientifically known to be the relevant means to felicity.

The third objection is best understood in light of Hobbes’s definition of linguistic reasoning as “nothing but *Reckoning* (that is, Adding and Subtracting) of the Consequences of generall names.”<sup>59</sup> This passage about *reasoning* is often taken to have implications for the kinds of normative *reasons* agents could have. It is sometimes thought that because Hobbes’s conception of reasoning is purely “formal,” he cannot acknowledge any non-instrumental normative reasons to desire some end or to act. I take this to be a non sequitur, especially in light of Hobbes’s foundationalism, but the thought is widespread enough to warrant consideration.<sup>60</sup> John Deigh, for example, asks whether on Hobbes’s view “reason determines the proper ends of action or is restricted to determining the best

<sup>54</sup> L 7.3–4: 98; 46.1–2: 1052.    <sup>55</sup> L 15.40: 242.

<sup>56</sup> Deigh (1996). Cf. Watkins (1973: 63, note 47).    <sup>57</sup> L 14.3–4: 198–200.    <sup>58</sup> L 15.40: 242.

<sup>59</sup> L 5.2: 64.

<sup>60</sup> As Gert (2001: 244) observes, “the currently almost universally accepted” view is “that Hobbes holds some formal account of reason, . . . which does not place any limit on the content of rational action.”

means to ends set by the agent's desires." He replies, on the basis of Hobbes's definition, that "reason" cannot "determine" the proper ends of action. His argument is this: (a) the "operations" of reason, according to Hobbes, "are entirely formal"; (b) this implies that Hobbes "excludes material criteria from the concept of reason"; (c) it is only by appeal to "a material criterion of reason" that "reason by its very nature" could "determine the proper ends of action."<sup>61</sup> If sound, Deigh's argument would rule out the normativity of PP, which prescribe desires and actions directed to an end not wholly grounded in one's present desires or aims.

There are two potential sources of confusion here, and to avoid them we must distinguish two senses of 'reason' and 'determine'. If 'reason' in all three premises means the activity of reasoning, then it is true that, in one sense of 'determine', reason cannot determine the proper ends of action: Hobbes did not think that one's *reasoning* could *make* it the case that one has a normative reason to realize those ends. But it is a separate question which normative *reasons* one has, independently of reasoning: if 'determine' is supposed to mean discover, then Hobbes did think that reasoning could determine what normative reasons one has. This is obvious in the case of epistemic reasons: by engaging in theoretical reasoning, one lays out and represents to oneself in the imagination the evidence for and against some proposition, and thereby discovers whether one has sufficient normative reasons for believing it. The same is true for practical reasons: to say the law of nature is a "Precept of Reason" is to say it is discovered or "found out by Reason."<sup>62</sup>

There is of course a sense in which Hobbes's account is "formal": reasoning consists in *operating* on contents; it does not consist *in* contents. If this is what is meant by 'formal', and if by 'reason' is meant reasoning, then it is true, as the argument's first premise says, that Hobbes's conception of reason is formal. This is what Deigh must mean, since the warrant for this premise is supposed to be Hobbes's definition of *reasoning*. That reasoning is formal in the sense just specified, however, has no intrinsic and direct bearing on the content of normative *reasons*: it does not determine whether there are normative reasons for agents to desire some ends non-instrumentally. Deigh therefore faces a dilemma. On the one hand, if in the third premise 'reason' means reasoning and 'determine' means make, then the argument is viable, but has no direct bearing on whether there are

<sup>61</sup> Deigh (1996: 39–40). I reconstruct Deigh's argument by adding 'only' to the third premise, which the argument needs.

<sup>62</sup> L 15.8: 226; 14.3: 198.

non-instrumental reasons. On the other hand, I take Deigh to be trying to show that Hobbes's account of reason rules out non-instrumental *reasons*. That would require 'reason' in the third premise to mean 'normative reasons', which, since 'reason' in the first premise must mean 'reasoning', inflicts a fallacy of equivocation on the entire argument. Deigh's argument conflates reasoning with normative reasons (or with rationality as a normative source of reasons).

Sometimes Hobbes's conception of reasoning has been thought to be "formal" in the different sense that, *qua* activity, it has no intrinsic, substantive end. Yet the whole point of philosophical reasoning, for Hobbes, is to discover causes and effects and one's *normative* reasons to believe, desire, and act. For Hobbes, linguistic reasoning is rule governed: one is not engaged in reasoning unless one operates on semantic or mental contents by seeking correctly to apply the correct rules of addition and subtraction. One may therefore reason rightly or wrongly depending on whether one applies the correct rules in the correct way. There is a difference between the reasoning in which any individual actually engages and correct or right reasoning.<sup>63</sup> And "the Method of Reasoning" – the set of rules agents must seek to apply for their activity to count as reasoning, and which agents must correctly apply to reason rightly – comprises the rules of logic.<sup>64</sup> To be sure, some of these rules, such as non-contradiction, are purely formal: they provide criteria for making valid inferences from starting premises. But for Hobbes right reasoning also includes substantive norms, a point he made when enumerating the three types of "defect" or "Error" that afflict practical reasoning: people reason incorrectly not only by drawing "Erroneous Inferences from True Principles," but also when they draw (valid) inferences from "false Principles" or "grounds of their Reasoning" or when they "mis-interpret" the reasons of which they are aware.<sup>65</sup> Reasoning from unsound premises or principles counts as an error in *reasoning* precisely because reasoning has an end or *point*. Actual reasoning often fails in this function, but right reasoning, by definition, does not.

Therefore, Hobbes's equation of the "Method of Reasoning" with rules of "logic" does not imply reasoning is purely formal in the sense of not having an end. Logic for Hobbes was, as for others in his time, a more expansive notion. We find Hobbes's most extensive account of logic in the section of *De Corpore* titled "Computatio sive Logica." The section's [final chapter](#), "De Methodo," defines the method of philosophical reasoning as

<sup>63</sup> L 5.3: 66.    <sup>64</sup> L 46.II: 1060.    <sup>65</sup> L 27.10–12: 458–260. Cf. DC 5.2.

the “shortest way of finding out [*investigatio*] effects by their known causes, or of causes by their known effects.” One discovers effects from causes via the method of addition or composition (*compositio*) and causes from effects via the method of subtraction or resolution (*resolutio*).<sup>66</sup> These two components jointly amount to a *logic of discovery*, which Hobbes called the *methodo inveniendi*. Do not be misled by the connotations of ‘invention’ in English today: Hobbes was following a long tradition in which *inventio* meant discovery. To be sure, in classical and Renaissance theories of rhetoric, *inventio* denoted the discovery of rhetorically persuasive arguments, which the orator was typically advised to do by finding commonplaces and by favourably disposing his audience.<sup>67</sup> But by the sixteenth and seventeenth centuries, under the influence of Lorenzo Valla, Rudolph Agricola, and, above all, Petrus Ramus, many began to classify *inventio* as a part of the art of dialectic or logic rather than rhetoric. *Inventio* was transformed into the simple discovery of “reason, proof, argument.”<sup>68</sup> We know Hobbes was, in writing of *methodo inveniendi*, following this later appropriation of *inventio*: the phrase appears in a text explicitly devoted to logic.<sup>69</sup>

Hobbes had used the equivalent vocabulary in *Leviathan*, when he called the mental activity of “hunting out” causes and effects “*Seeking*,” which he equated to the exercise of one’s “faculty of Invention” (LL: “*Investigationem, sive facultatem Inveniendi*”).<sup>70</sup> Seeking or discovering causes is a substantive purpose of philosophical reasoning.<sup>71</sup> The ultimate “Use and End of Reason” is not just to find the “truth of one, or a few consequences” drawn from uncertain premises, but to begin from self-evident first principles known with “certainty” and embodied in “first definitions, and settled significations of names,” and then to “proceed from one consequence to another,” so that the “certainty of the last Conclusion” of one’s reasoning is inherited from the “certainty of all those Affirmations and Negations, on which it was grounded.”<sup>72</sup>

<sup>66</sup> DC 5.1. <sup>67</sup> Skinner (1996: 45–46, 259, 360).

<sup>68</sup> Skinner (1996: 58–59), quoting Ramus’s 1555 French digest of his 1543 *Dialecticae Partitiones*.

<sup>69</sup> We also know this because Hobbes contrasted the method of invention to the method of teaching (*methodo docendi*). DC 6.10. While Skinner notes that Hobbes was often hostile to *inventio* as understood in the rhetorical tradition, he fails to note that Hobbes appropriated *inventio* from this Ramist tradition for his account of reasoning.

<sup>70</sup> L 3.5: 42; LL 3: 43.

<sup>71</sup> L 46.1: 1052; DC 1.2. Pace Deigh (1996), philosophical reasoning does not exclude, but *improves* on discovering causes. L 3.11: 46. Hoekstra (2003).

<sup>72</sup> L 5.4: 66.

Nor is the end of reasoning merely to “invent” or discover causal relations – to discover, for example, that  $\phi$ -ing is the relevant means to something. It is also to discover that one has epistemic reasons, for example to believe that  $\phi$ -ing is the relevant means and, indeed, to discover *practical reasons* to act. When Hobbes deplored the fact that people “by mis-reasoning . . . fall upon false and absurd generall rules,” he was in part writing of reasoning about the “rules of their actions,” and when he adumbrated the “defects in Reasoning,” he specifically mentioned reasoning that ends in “concluding, and resolving what to do.”<sup>73</sup> We reason about what to do.

A critic might object that the point of Hobbesian reasoning is not to discover normative reasons, but simply to discover causal relations; such discoveries consequently move us to take the means to our desire’s object. Reasoning does not discover reasons, on this view, because its initial starting point or “principle” is not a normative premise (such as the prudential prescription to care for one’s felicity), but a psychological fact (that we do care for our felicity or self-preservation). The critic might point to the following example of Hobbesian practical reasoning:

If a man by the terrour of present death, be compelled to doe a fact against the Law, he is totally Excused; because no Law can oblige a man to abandon his own preservation. And supposing such a Law were obligatory; yet a man would reason thus, *If I doe it not, I die presently; if I doe it, I die afterwards; therefore by doing it, there is time of life gained*; Nature therefore compells him to the fact.<sup>74</sup>

The objection is unfounded, however. On Hobbes’s weak psychological egoist-hedonist view, intentional agents are not motivated to act unless they see the object of their action as pleasing to and hence good for them, but they may very well fail to act for their own true felicity. They may reason improperly and be motivated by “irrationall” appetites that compromise their long-run good or self-preservation; indeed, they may be Cyrenaics who mistakenly believe their ultimate good lies in immediate pleasures. Yet even for these agents, the prudential precepts remain normative: it is not just “Nature,” but “Reason, which dictateth to every man his own good.”<sup>75</sup>

The critic might concede that one point of reasoning is to discover normative reasons, but nevertheless insist that reasoning can only do so on

<sup>73</sup> L 5,18–19: 74; 27,12: 460.

<sup>74</sup> L 27,25: 468. For such a non-normative, psychological reading, see Olsthoorn (2014).

<sup>75</sup> L 15,4: 222. See Darwall (2000: 318): “according to Hobbes, we reason *from the end* of self-preservation itself, not from the fact that it is our end.”

the basis of taken-for-granted ultimate ends (such as felicity) or foundationally normative precepts (such as PP) whose normativity is not itself discoverable via reasoning. By reasoning one can discover causal relations – for example, that an illegal action would forestall death – and thereby discover practical reasons *given* the normativity of an ultimate end or foundationally normative precept; but one could not via reasoning discover the normativity of such an end or precept.

There is much truth in this, but only if by “discovered via reasoning” is meant *inferentially discovered* or *deduced*. It is true that we cannot rationally *deduce* or inferentially demonstrate the normativity of ultimate ends or of foundationally normative precepts. But right reasoning for Hobbes includes adherence not just to formal rules of logical deduction or demonstration, but also to substantive and self-evident or “manifest” truths. Scientific reasoning is about causes for Hobbes, but science must ultimately be grounded in foundational principles that are self-evidently or manifestly true in virtue of the fact that rational agents on reflection could not but *conceive* them to be true. This conceivability or intelligibility standard is at play in our knowledge of the fundamental ontological truths of Hobbes’s corpuscularian materialism, according to which all effects are caused by bodies in motion, truths that “are manifest of themselves [*manifestae sunt per se*], or (as they say commonly) known to nature [*naturae (ut dicunt) nota*]; so that they need no method [i.e., analysis] at all.”<sup>76</sup> Such “universal principles . . . cannot be demonstrated; and seeing they are known by nature [*sint nota naturae*] . . . need no demonstration, though they need explication.”<sup>77</sup> We do not discover these fundamental ontological truths via syllogistic reasoning; they are “manifest” or self-evident to us by “nature” in virtue of our inability to conceive of them as not true. This is why Hobbes defined a cause – the fundamental notion of his mechanistic determinism – in terms of the conceivability standard: “a cause is the sum or aggregate of all such accidents, both in the agents and in the patient, as concur to the producing of the effect propounded; all which existing together, it cannot be understood [*intelligi non potest*] but that the effect existeth with them; or that it can possibly exist if any one of them be absent.”<sup>78</sup> Like Aquinas, Hobbes was a foundationalist.

Thus when Hobbes wrote, in the passage quoted previously, that practical reasoners would be compelled by “Nature” to the illegal action, the term does not merely denote what they would actually be compelled to do. It also denotes what they would self-evidently know they ought to do –

<sup>76</sup> DC 6.5.    <sup>77</sup> DC 6.12.    <sup>78</sup> DC 6.10.

what is “known to nature” – without analytical ratiocination. Hobbes deployed the conceivability standard not only for knowledge of the mechanical causes of sensory perception, but also for our implicit knowledge of foundationally normative precepts. I may be able to discover that peace is a relevant means to self-preservation via reasoning that draws on causal information, but my knowledge of the foundationally normative Prudential Precepts depends on the non-empirical, non-inferential insight that “Reason is the *pace*; Encrease of Science, the *way*; and the Benefit of mankind, the *end*.”<sup>79</sup> This is why in *De Cive* Hobbes characterized aversion to death not just as one of the certain postulates of human nature (*duo certissima naturæ humane postulata*), but in *normative* terms as a postulate of *natural reasoning* (*rationis naturalis*) by which each strives to avoid violent death as the supreme natural evil (*summum naturæ malum*).<sup>80</sup>

### 3.4 Cognitive Subjectivism about Reasons? Epistemically Relativized Objective Reasons

The fact that the standard of *right* reasoning does not constitute agents’ normative reasons leaves open the possibility that their reasons are partly determined by their *actual* reasoning. This possibility is suggested by the fact that Hobbes often used the term ‘reason’ to mean agents’ own *act* of reasoning and/or the subjective *judgement* about reasons they form upon reasoning. Hobbes usually signalled this use of ‘reason’ by using possessives or by coupling ‘reason’ with terms such as ‘judgement’ and ‘conscience’. Chapter 26 of *Leviathan* furnishes an example: when Hobbes wrote that “in all kinds of actions, by the laws praetermitted, men have the Liberty, of doing what their own reasons shall suggest, for the most profitable to themselves,”<sup>81</sup> he meant that, where the law is silent, subjects have the liberty of doing what in their own reasoned judgement they deem most profitable. This is an especially clear example because the passage deploys ‘reason’ as a count noun, in the plural. Other passages are less conspicuous, because when Hobbes meant by ‘reason’ an agent’s subjective, reasoned judgement, he almost always deployed the term as a mass noun, as when he wrote that “Impediments . . . cannot hinder him from using the power left him, according as his judgement, and reason shall dictate to him,”<sup>82</sup> or when he defined an agent’s right of nature as the liberty “of doing any thing, which in his own Judgement, and Reason, hee shall conceive to be the aptest means” of self-preservation.<sup>83</sup>

<sup>79</sup> L 5.18: 74.    <sup>80</sup> DCv Dedicatory.10.    <sup>81</sup> L 21.6: 328.    <sup>82</sup> L 14.2: 198.    <sup>83</sup> L 14.1: 198.

These uses of ‘reason’ are especially important for Hobbes’s political philosophy. The issue is whether one’s own private reasoned judgement is *normative* for oneself, i.e., whether one’s judgement that one has certain reasons could itself also *make* it the case that one does have such reasons.<sup>84</sup> If yes, then although Hobbes was not committed to *conative* subjectivism, he would have been committed to *cognitive* subjectivism about reasons.

Several important passages seem to support a cognitively subjectivist reading. First, in *Leviathan* 27.3 Hobbes wrote that, with respect to the precepts of natural law, each is “cleared by the Uprightnesse of his own Intention. When therefore his Intention is Right, his fact [i.e., action] is no Sinne: if otherwise, his fact is Sinne.”<sup>85</sup> It would therefore seem that if one thinks that one has not violated natural law, then one has not done so. Second, in *Leviathan* 29.7 Hobbes wrote: “he that is subject to no Civill Law, sinneth in all he does against his Conscience, because he has no other rule to follow but his own reason.”<sup>86</sup> It would therefore also seem that if one thinks one has committed a violation, then one has done so. The appeal to right intention or to conscience in each case appears to suggest that the normative reasons one has are relativized to the reasons one *thinks* one has.

To see why the appearance is deceiving, we must distinguish two ways in which subjective states might be implicated by a precept. A precept might prescribe  $\phi$ -ing just in case one thinks one has a reason to  $\phi$ . This is a true instance of a cognitively subjective precept: its prescription is relativized to the reasons one *thinks* one has. By contrast, a precept might prescribe a certain subjective state, such as a desire or belief. This is not a subjective precept in the relevant sense: its *object* is a subjective state, but it may well be an objective desire-precept.

A comparison of *Leviathan* 27.3 to 15.36 in light of this distinction suggests that the second, non-subjective interpretation is correct. The latter, well-known but difficult, passage is introduced with the marginal heading “The Lawes of Nature oblige in Conscience alwayes, but in Effect then onely where there is Security.” The passage reads:

The Lawes of Nature oblige *in foro interno*; that is to say, they bind to a desire they should take place: but *in foro externo*; that is, to the putting them in act, not alwayes. For he that should be modest, and tractable, and

<sup>84</sup> Skorupski (2007: 83–84) observes that Williams’s (1981) existence-internalism entails such bootstrapping.

<sup>85</sup> L 27.3: 454.

<sup>86</sup> L 29.7: 502. “It is plain, that to do what a man thinks in his own conscience to be sin, is sin; for it is contempt of the law itself.” AB 376. See EL 25.12.

performe all he promises, in such time, and place, where no man els should do so, should but make himselfe a prey to others, and procure his own certain ruine, contrary to the ground of all Lawes of Nature, which tend to Natures preservation.<sup>87</sup>

The correct interpretation of this passage focusses on the distinction between affective reasons for desire or disposition and practical reasons for action. Hobbes elsewhere characterized the laws of nature as intrinsically dictating “Vertues,” “habits of the mind,” or “qualities that dispose men to peace.”<sup>88</sup> In this passage he was making the same point, claiming that natural laws articulate decisive affective reasons, binding under all circumstances, to “desire,” “endeavour,” or be *disposed* to act peacefully, but that they provide decisive practical reasons to *act* peacefully only when it is safe to do so. As Hobbes put it three paragraphs later, the laws of nature “require nothing but endeavour; he that endeavoureth their performance, fulfilleth them.”<sup>89</sup> Invoking the inner court of “Conscience” here signals the fact that the primary *object* of the prescription is a subjective state, not that the prescription is cognitively subjective and so relativized to what the individual *thinks* is normative. For a precept to “oblige in Conscience” is therefore for it normatively to prescribe a subjective state consisting in a disposition. *Leviathan* 27.3 should be read in the same way: to have the right “Intention” is akin to having the right desire or disposition. If one is disposed to act peacefully, but fails to act peacefully because doing so is unsafe, one still fulfils the law of nature, whereas if one is not disposed to act peacefully, then one violates the law of nature however one acts.

The interpretation I have offered faces two problems, however. The first arises from another cognitively subjective-sounding passage, *Leviathan* 15.37: “whatsoever Lawes bind *in foro interno*, may be broken, not onely by a fact [i.e., action] contrary to the Law, but also by a fact according to it, in case a man think it contrary. For though his Action in this case, be according to the Law; yet his Purpose was against the Law; which where the Obligation is *in foro interno*, is a breach.”<sup>90</sup> This makes it appear that “oblige in Conscience” is not (only) normatively to prescribe a virtuous subjective disposition, but (also) normatively to prescribe a subjective disposition or action that one subjectively *thinks* is virtuous. The problem’s source is this: Hobbes’s invocation of the inner court of “Conscience” seems to equivocate between denoting one’s *conative* attitudes versus denoting the *relationship* between one’s conative attitudes or

<sup>87</sup> L 15.36: 240.      <sup>88</sup> L 26.36: 442; 26.8: 418. See DCv 3.29; Boonin-Vail (1994).      <sup>89</sup> L 15.39: 240.  
<sup>90</sup> L 15.37: 240.

actions to *cognitive* attitudes such as beliefs (i.e., denoting whether one's desires or actions are as one thinks they should be).

The second problem is that whereas my interpretation makes perfect sense of *Leviathan* 27.3, applying it to *Leviathan* 29.7 runs against the grain of Hobbes's wording: like cognitively subjective views, writing that a person "sinneth in all he does against his Conscience"<sup>91</sup> invokes a *relationship*, between one's *action* ("all he does") and the content of one's conscience. Two passages – 15.37 and 29.7 – therefore provide prima facie evidence for a cognitively subjective interpretation.

But to see Hobbes's rejection of the cognitively subjectivist view, consider someone who misunderstands the non-social laws of nature, temperance for example.<sup>92</sup> Even if someone subjectively *thinks*, contrary to fact and the epistemically accessible evidence, that abusing alcohol will do his health no harm, this would not annul the normative reason he has to (be disposed to) refrain from alcohol abuse. The precept of temperance, like all natural laws, articulates a reason for agents not relativized to what they happen to think, but to the *epistemically accessible* evidence, i.e., evidence of which they could become aware via reflection or further (spatio-temporally possible) investigation or observation.<sup>93</sup> Hobbes explicitly denied that those who misunderstand or are unaware of the normative reasons articulated in natural laws are thereby excused: although many misinterpret the laws of nature, no "man, that pretendeth to reason" shall "be supposed to want means to know the Lawes of Nature; because they are known by the reason he pretends to: only Children, and Madmen are Excused from offences against the Law Naturall."<sup>94</sup> Children and madmen are excused because natural laws are not epistemically accessible to them. Those who could know the laws of nature, but do not, "need to be diligently, and truly taught."<sup>95</sup> For Hobbes, normative affective and practical reasons are relativized not to what agents actually believe, but to what they have a sufficient epistemic reason to believe.

The same objective but epistemically constrained conditions operate within particular laws of nature themselves. When Hobbes argued that, under the third law of nature prescribing faithfulness to covenants, any covenant of mutual trust undertaken "in the condition of meer Nature" is, "upon any reasonable suspicion" that the other party will renege, "Voyd," he did not mean that the covenant is void if one actually *thinks* the other will renege. It is void if one has a sufficient

<sup>91</sup> L 29.7: 502.    <sup>92</sup> DCv 3.32; L 15.34: 238; 31.40: 572.    <sup>93</sup> Skorupski (2010: 41–43).

<sup>94</sup> L 27.23: 468; see 26.21: 430; R&C.13: 1139.    <sup>95</sup> L 30.4: 522.

reason to think the other will renege. The “suspicion” must be *reasonable*.<sup>96</sup>

Hobbes’s effort in *Leviathan* to define ‘conscience’ reflects his antipathy to cognitive subjectivism. He there began by arguing that the term was initially associated with awareness of criminal deeds,<sup>97</sup> which is why it became “reputed a very Evill act, for any man to speak against his *Conscience*,” i.e., it came to be considered evil to lie about crimes. Hobbes then proceeded to cast aspersions on views that ground agents’ normative reasons in their subjective beliefs: in subsequent generations, “men, vehemently in love with their own new opinions, (though never so absurd,) and obstinately bent to maintain them, gave those their opinions also that revered name of *Conscience*, as if they would have it seem unlawful, to change or speak against them; and so pretend to know they are true, when they know at most but that they think so.”<sup>98</sup>

The question therefore is what aspects of Hobbes’s thought are being expressed in *Leviathan* 15.37 and 29.7 about the normative significance of the conscience. The answer is that the passages reflect Hobbes’s view that when agents do not do (or are not disposed to do) what they judge they ought to, they violate one of the reasons they have and so are open to criticism. It does not follow, however, that they have a reason or ought to (be disposed to) do whatever they judge they have a reason or ought to do: agents might judge that they ought (to be disposed) to do some very imprudent or wicked things, in violation of natural law. The two passages do not express the view that, if one judges that one ought to  $\phi$ , then one ought to  $\phi$ . Rather, they express the view that one has a reason not to *both* fail to  $\phi$  and judge that one ought to  $\phi$ . The passages express a wide-scope reason not to combine the attitude and action or disposition, rather than a narrow-scope reason for the action or disposition.<sup>99</sup> Consider the agent who judges, contrary to the seventh natural law, that he ought to be disposed to wanton cruelty,<sup>100</sup> and yet, because of fear or cowardice, fails to be so disposed. His disposition conforms to the law, and yet is “against his *Conscience*.” He is therefore criticizable for violating the wide-scope reason he has not to combine the disposition with the judgement; but since

<sup>96</sup> L 14.18: 210.

<sup>97</sup> The term Hobbes used was “fact,” which according to *OED* most commonly meant an evil deed or crime. Commentaries on this paragraph widely fail to take this meaning into account (e.g. Tralau (2011) and the cited literature).

<sup>98</sup> L 7.4: 100.

<sup>99</sup> Broome (2007b). The wide-scope reason is akin to “enkrasia” (Broome 2013: 170).

<sup>100</sup> L 15.19: 232.

he ought to have the disposition he actually has, it follows that he has a reason not to judge as he actually judges – not that he ought to have the disposition he judges he ought to have. The violation of natural law stems from the judgement or belief, not the disposition.<sup>101</sup>

*Leviathan* 29.7 also reflects a further commitment: that rational agents in the state of nature have a reason to be disposed to do whatever they themselves judge, via their own reasoning, that they have decisive reasons to do.<sup>102</sup> There is therefore a weak sense in which one's own reasoned judgement is normative in the state of nature: one does not have a reason to do whatever one happens to judge one ought to do, but to be *disposed* to do what one judges one ought to do as a result of *reasoning*. This is because, first, reasoning is the most reliable available means for discovering one's normative reasons; second, one has a reason to prefer one's own reasoned judgements, because all are biased in their own favour;<sup>103</sup> and third, by right of nature one does not owe it to anyone to defer to anyone else's reasoned judgement.<sup>104</sup> This is why, as Hobbes put it in that passage, each in the state of nature "has no other rule to follow but his own reason."<sup>105</sup> But from the fact that one has a reason or ought to be *disposed in general* to act on one's own reasoned judgement, it does not follow that one ought to do whatever one judges one ought to do *in each case*: it is not true that one ought to do what one mistakenly judges one ought to do.<sup>106</sup>

Hobbes's commitment to an objective but epistemically constrained theory of affective and practical reasons is also reflected in his treatment of the means that count as *relevant* to one's normative ends, i.e., the means to which the normativity of ends is transmitted. There are at least two questions we might ask about how to interpret relevance. The first is whether the relevant means consist in cognition-independent, *objective* conditions, or in cognition-dependent, *subjective* conditions. On one cognitively objective interpretation, relevant means are those that *actually* promote the end; on a typical cognitively subjective interpretation, they are what the agent *thinks* promote the end. Hobbes's definitions of natural law in *Elements* and *De Cive* use objective-sounding language, but the English *Leviathan* equivocates between objective- and subjective-sounding language: a law of nature is there said to be "a Precept, or generall Rule, found out by Reason, by which a man is forbidden to do, that, which is destructive of his life, or taketh away the means of preserving the same; and to omit, that, by which he thinketh it may be best

<sup>101</sup> See Boonin-Vail (1994: 88, 112, 183–186). <sup>102</sup> DCv 2.1, note. <sup>103</sup> L 26.21: 430.

<sup>104</sup> EL 14.8; DCv 1.9; L 14.1: 198. <sup>105</sup> L 29.7: 502. <sup>106</sup> Broome (2013: 199–201).

preserved.”<sup>107</sup> This equivocation has rightly puzzled readers attending to it; recognizing that Hobbes rejected the cognitively subjective interpretation of natural law, some have suggested the subjective-sounding language is best treated as an oversight.<sup>108</sup> That it is no mere oversight, however, is indicated by the fact that when Hobbes condensed this passage – as was his usual practice – in his Latin translation of *Leviathan* (1668), by dropping the distinction between commission and omission and hence leaving room for only one formulation, he retained the cognitively subjective-sounding language.<sup>109</sup>

These equivocations are not a mere sign of carelessness: they are a sign of a genuine, but not always fully articulate, feature of Hobbes’s thought. What ultimately explains the equivocal language is that his considered view lies between the objective “true-means” and cognitively subjective “thought-to-be-means” interpretation of natural law and the principles in which they are grounded. As I have suggested, Hobbes defended an *objective*, but *epistemically constrained* conception of ITP and PP. He most clearly articulated this conception in his well-known reply to the Foole in chapter 15 of *Leviathan*. The Foole rejects the normativity of the third law of nature, which proscribes the violation of covenants, by claiming as normative both true-means and thought-to-be-means versions of Instrumental Transmission and Prudence. Taking self-preservation and contentment to be normative ends, the Foole claims that (a) rational agents each ought to take the *actually* best means to them (“those actions are most Reasonable, that conduce most to their ends”), from which it follows that none has a decisive reason not to take those means; he further claims that (b) “there could be no reason, why every man might not do what he thought conduced” to his “conservation, and contentment.” He concludes that to have broken one’s covenants “was not against Reason, when it conduced to ones benefit.”<sup>110</sup>

Although the Foole’s conclusion appeals only to the former, objective version of ITP and PP (he proceeds to ignore the latter, cognitively subjective version), Hobbes’s response to the Foole involves rejecting both the objective, true-means and the subjective, thought-to-be-means versions. Hobbes asserted instead that one has a reason to take the means that one can reasonably *know* (via science) or *foresee* (via epistemic prudence) would promote one’s normative ends.<sup>111</sup> One can *reasonably* know

<sup>107</sup> EL 15.1; DCv 2.1; L 14.3: 198.      <sup>108</sup> Kavka (1986: 340–341); cf. Hoekstra (1997: 630).

<sup>109</sup> LL 14: 199.      <sup>110</sup> L 15.4: 222.

<sup>111</sup> On scientific knowledge versus expectation and epistemic prudence, see L 3.7–9: 42–44; 5.17–22: 72–76; 46.1–2: 1052; DC 5; de Jong (1986); Jesseph (1996, 2010).

or foresee something in the sense that there is sufficient cognitively accessible evidence for it, such that one has a sufficient epistemic *reason* to know or believe it. Thus Hobbes argued that “when a man doth a thing, which notwithstanding any thing can be foreseen, and reckoned on, tendeth to his own destruction, howsoever some accident which he could not expect, arriving may turne it to his benefit; yet such events do not make it reasonably or wisely done.”<sup>112</sup> One does not have a reason to take the actually true means (perhaps only discoverable *ex post* with hindsight) to one’s good; nor does one have a reason to take what one happens to think (perhaps erroneously) are the means; rather, one has a reason to take what one could reasonably know or foresee, *ex ante*, based on the reasons for belief provided by the epistemically accessible evidence, to be the means.<sup>113</sup>

The Foole, by contrast, declares that if “by Rebellion” one will actually end up “attaining Sovereignty,” then breaking one’s covenant to obey the sovereign “stands with . . . Reason.” Hobbes’s retort accordingly appeals to his epistemically constrained version of ITP and PP: “it is manifest, that though the event follow, yet because it cannot reasonably be expected, but rather the contrary . . . the attempt thereof is against reason.”<sup>114</sup> Similarly, concerning state-of-nature covenants in which one party has already performed his part, the Foole declares that sometimes it is “against reason, that is, against the benefit of the other to performe.” Hobbes replied that no one could “hope” to preserve himself in the state of nature “without the help of Confederates,” and that a reputed covenant breaker “cannot be received into any Society . . . but by the error of them that receive him . . . which errors a man cannot reasonably reckon upon as the means of his security.” An outcome – even a greatly advantageous outcome – that one “could not foresee, nor reckon upon” does not provide one with any reason for action: rather, one has a reason to take the means that one could reasonably know or foresee would promote one’s normative ends.<sup>115</sup> On Hobbes’s account, facts that rational agents could not reasonably know or foresee – given the

<sup>112</sup> L 15.5: 224.

<sup>113</sup> Cf. Parfit (2011) on what one ought to do in the fact-, belief-, and evidence-relative senses, and on what is “expectably best.” The Foole interprets reasons in a fact-relative way, Hobbes in an evidence-relative way. The laws of nature are “theorems, tending to peace, and those uncertain, as being but conclusions of particular men” (AB: 285).

<sup>114</sup> L 15.7: 224; 15.4: 222; 15.7: 224. The Latin version: “even if it is so attained, it will have been attained against right reason . . . because the success of such enterprises is uncertain from the beginning.” LL 15: 225.

<sup>115</sup> L 15.5: 224. For a similar interpretation, see Darwall (1995: 73–77). Hoekstra (1997: 621, 642 note 5) argues against this interpretation on the grounds that it lies “in evident tension with Hobbes’s” supposed view, attributed by Hampton (1986: 65), that parties to state-of-nature covenants face a prisoner’s dilemma. I argue against Hampton on war in Abizadeh (2011).

epistemically accessible evidence – cannot provide them with reasons for action because such reasons are supposed to *guide* agents and to provide grounds for normative criticism of them. One cannot be normatively blamed for failing to heed inaccessible considerations. Something that could be known or observed only *after the fact* could very well provide an *ex post* reason to regret not having taken what turned out to be the true means to one's normative ends, but what is unknowable and unforeseeable beforehand cannot furnish an *ex ante* affective or practical reason. It is both reasonable not to take means that would turn out, by “accident” in defiance of the odds, to realize one's end, and reasonable afterwards to regret not having done so.

The reply to the Foole also clarifies Hobbes's response to a second question about the relevance of means: In what sort of *promoting relation* do relevant means stand to their ends, i.e., what does it mean for something to *promote* an end? Are the relevant means conditions *necessary* for realizing the end, *sufficient* for realizing the end, the most *efficient* or cost-effective means for realizing the end, or dispositionally render the end's realization more *likely*? Hobbes's language is equivocal, but the overall picture that emerges from his various formulations is dispositional. In *Elements*, he wrote of the “Lawes of nature that tend to our preservation” and characterized them as “precepts that tend to peace.”<sup>116</sup> It is true that in *De Cive* he usually characterized the means specified by natural law as what is “necessary to obtain peace [*conciliandæ paci necessariam*].”<sup>117</sup> But as we have seen, natural-law theorists often used ‘necessary’ loosely – as Grotius had put it, “not with physical exactitude, but in a moral sense” – not to mean indispensable, but roughly *conducive* to realizing an end.<sup>118</sup> Hobbes followed this usage in his characterization of the right of nature, which he typically characterized as a right to the (subjectively) necessary means to self-preservation.<sup>119</sup> The nineteenth chapter of *Elements*, for example, glosses an individual's right to “the necessary means of his preservation” as equivalent to a right to what “tendeth to his own Safety or Commodity.”<sup>120</sup> Similarly, the tenth paragraph of *De Cive*'s opening chapter treats it as a right to those means “that necessarily conduce [*quæ necessariò conducunt*] to the protection of one's life.”<sup>121</sup> Hobbes used the

<sup>116</sup> EL 17.14; 18.3.   <sup>117</sup> DCv 3.1; see 3.13; 3.19; 3.20; 3.31.   <sup>118</sup> JBP III.1.2.

<sup>119</sup> Writing of “the necessity of the meanes” in *Elements*, Hobbes argued that “it is not against Reason and therefore Right for a man to use all meanes and doe whatsoever action is necessary for the preservation of his body.” EL 14.7–8. See EL 24.2 (“the onely meanes, by which we have any possibility of preserving ourselves”) and DCv 1.8; L 15.17: 232; 28.2: 482.

<sup>120</sup> EL 19.1; see 14.10.   <sup>121</sup> DCv 1.10; cf. 1.9.

same language of necessity to denote dispositional conditions in his characterization of natural laws, as when he wrote that they dictate “all things necessarily conducting to Peace [*ad pacem necessariò conducunt*].”<sup>122</sup> While sometimes the English *Leviathan* characterizes the laws of nature in terms of “their necessity, for the procuring, and maintaining peace,”<sup>123</sup> Hobbes there usually wrote (and in definitional contexts always wrote) of the “best” or “the aptest means” to self-preservation – which terms he translated with the dispositional verb *tendere* in the Latin *Leviathan*.<sup>124</sup> This is also how he characterized the relevant means prescribed by reason in his reply to the Foole: in terms of what “tendeth to his own destruction.”

The traditional (Aristotelian) way to understand the notion of a disposition or tendency was qualitatively, as equivalent to a *potential* that will actualize itself under appropriate, qualitatively specifiable conditions. But Ian Hacking has argued that seventeenth-century Europe saw the public emergence of the modern concept of *probability*, which, he argues, is a Janus-faced concept combining two aspects: an *aleatory* or statistical aspect, concerned with the frequency with which particular types of event occur over repeated trials; and an *epistemological* aspect, concerned with the degree of confidence in beliefs as warranted by the evidence. The first, aleatory aspect was “connected with the tendency, displayed by some chance devices, to produce stable relative frequencies.”<sup>125</sup> Probability thus furnished a different way to understand tendency, quantitatively, in terms of *likelihood* or the *frequency* with which some disposition will manifest itself over multiple trials and our *confidence* in predicting such outcomes.

I take it that what explains Hobbes’s equivocal language, when describing relevant means, is that he was circling around this newly emergent notion of probability. On the best reading of Hobbes, the relevance of means is a function of their capacity to increase the *probability* of realizing ends, where probability is understood in terms of likelihood given the epistemically available evidence. (As Hobbes’s discussion of “contingency” indicates, this notion of probability is compatible with his determinism.<sup>126</sup>) True, Hacking suggests the concept of quantifiable probabilities emerged publicly in Europe only around 1660 – i.e., after the completion of Hobbes’s three political treatises. But Richard Tuck has provided evidence of its public emergence by name in English texts dating to the 1630s from the Tew Circle, whose members met at Lord Falkland’s house in

<sup>122</sup> DCv 9.14. See the use of ‘tendit’ in DCv 3.32.   <sup>123</sup> L 26.22: 430.

<sup>124</sup> L 14.3: 198; 14.1: 198; LL 14: 199.   <sup>125</sup> Hacking (2006: 1).   <sup>126</sup> Q: 37/49–50.

Oxfordshire in that decade. As Tuck notes, the problem of decision-making under risk and uncertainty was a live problem for these thinkers, “and they were often drawn to a solution which depended upon the estimation of probabilities.”<sup>127</sup>

It might be objected that Hobbes himself did not provide any quantitative analysis of the probabilities involved in determining what to do – as the Port-Royal *La Logique* did a few years later in 1662.<sup>128</sup> But, as Hacking acknowledges, as early as 1640 Hobbes did articulate the emergent notion of measurable probabilities in his discussion of signs. Having explained that observers take something to be a sign of another when the two, like clouds and rain, have been previously observed in constant conjunction as antecedent and consequent, Hobbes went on to remark that signs taken from experience

are but Conjecturall and according as they have often or seldome failed, so theire assurance is more or lesse; but never full and evident. For though a man hath alwayes seene the day and night to follow one another hitherto, yet can he not from thence conclude they shall doe soe, or that they have done soe eternally. Experience concludeth nothing universally. If the Signes hit twenty tymes for once missinge, a man may lay a wager of twenty to one of the event, but may not conclude it for a trueth.<sup>129</sup>

This text combines an epistemological focus on degrees of belief (“assurance . . . more or lesse”) with an aleatory focus on frequencies over multiple trials. Hacking rightly notes that here the modern notion of “probability has emerged in all but name.”<sup>130</sup>

What Hacking does not note is that Hobbes explicitly linked this discussion of probabilities to *prudence*. The quoted text immediately continues:

But by this it is plaine that they shall conjecture best that have most experience, because they have most Signes to conjecture by; which is the reason that ould Men are more prudent, that is, conjecture better *Caeteris paribus*, than younge for being oulder they remember more, and experience, is but remembrance. And men of quicke Imagination *Caeteris paribus*, are more prudent than those whose Imaginations are slowe; fore they observe more in lesse tyme. And PRUDENCE is nothing else but Conjecture from Experience, or taking of Signes from experience warily . . . As in Conjecturall things concerning Past & Future. It is prudence to conclude from experience, what is likely to come to passe, or to have passed allready.<sup>131</sup>

<sup>127</sup> Tuck (1979: 104–105).      <sup>128</sup> Hacking (2006: 77–78).

<sup>129</sup> EL 4.10. See the discussion of likelihoods in L 5.5: 68.

<sup>130</sup> Hacking (2006: 48).

<sup>131</sup> EL 4.10–11.

When discussing sedition, Hobbes explicitly linked prudent, reasonable action to a measurable calculation of the odds. Any “Authour of Sedition” who considers “his experiences aright” concerning past seditions “shall finde, that for one man, that hath thereby advanced himselfe to honour, twenty have come to a reproachfull end.” Hobbes concluded that anyone faced with these stakes and who gambles on odds of twenty to one “must not be Prudent.”<sup>132</sup>

Both Hobbes’s account of the prudential *normative* reasons agents have (which help determine what they ought to do), and his account of the *motivational* reasons for action they have (which help explain what they in fact do), draw on this notion of the prudential estimation of probabilities. The former informs Hobbes’s response to the Foole, which combines the notions of prudential foresight, tendency, accidental odds, expected benefit, and normative reasons for action. The latter is illustrated by remarks elsewhere in *Leviathan* about subjects condemned to death by their sovereign, who can be expected to fight officers coming to seize them: for “man by nature chooseth the lesser evill, which is danger of death in resisting [the executioner]; rather than the greater, which is certain and present death in not resisting.”<sup>133</sup> The notion of *danger* here refers to the probabilistic risk of death, which yields a “lesser” expected evil than the “greater” one of *certain* death. It is true that Hobbes never actually quantified agents’ expected overall well-being. He was nevertheless thinking in quantifiable terms – both with respect to the amount of good at stake and the probability of realizing it – and this when thinking about prudential reasons.<sup>134</sup> In this Hobbes was no different than Grotius, who, a few decades earlier, had listed quantitative “rules of prudence” for deliberation, including: if the good and evil consequences at stake seem equal in amount, then act only if the “efficacy” for good (i.e., the probability of realizing it) are greater than for evil; more generally, act only if the ratio of (the amount of) good to evil is greater than the ratio of good’s efficacy to evil’s efficacy.<sup>135</sup>

On the best reading of Hobbes, the *relevant* means – the so-called “necessary” means – are those that make realizing the end in question more likely, relative to the epistemically accessible evidence. This newly emergent, not always fully articulated, and epistemically relativized notion

<sup>132</sup> EL 27.13.    <sup>133</sup> L 14.29; 214. Cf. LN §6.

<sup>134</sup> Tuck (1979: 128) denies the final point, but even if he is right that humans’ motivational reason for entering a commonwealth is grounded in certain, scientific knowledge, it does not follow that they do not draw on probabilistic, prudential “knowledge of fact” in deciding what to do. L 7.3–4: 98; 46.1–2: 1052.

<sup>135</sup> JBP II.24.5; cf. III.1.4.

of probability is what led Hobbes to alternate, in his mature formulations, between invoking the “best” means and those that “tend” to realize an end.

I conclude that ITP, PP, and the natural laws grounded in them are, like affective and practical reasons in general, relativized for Hobbes to the epistemically accessible evidence, i.e., to what one has sufficient reasons to believe – rather than to what one subjectively *thinks*, or to what actually turns out to *be* the case. Although many actually misinterpret the laws of nature, knowing them is nevertheless “easy to such, as without partiality, and passion, make use of their naturall reason, and therefore leaves the violators thereof without excuse.”<sup>136</sup>

### 3.5 Suicide, Laws of Nature, and a Life Worth Living: Self-Preservation Is Not Survival

According to the fourth objection against the prudentialist interpretation, prudentialism is no better than instrumentalism at explaining the universal normativity of self-preservation and, consequently, of natural law. Recall that a difficulty confronting the instrumentalist interpretation is that survival is not always a means for fulfilling agents’ present desires or aversions: fulfilling one’s *post-mortem* passions presupposes one’s death; and fulfilling *self-destructive* passions involves or would likely lead to one’s death. A similar difficulty seems to confront prudentialism: because survival is not always a means for fulfilling one’s strongest passions, the course of action that would enhance one’s overall well-being might usher in death. The prudentialist interpretation of Hobbes would then seem, like the instrumentalist interpretation, incapable of explaining how self-preservation and the laws of nature could be universally binding, including upon agents with (strong) post-mortem and self-destructive passions.

It might seem that one could answer this objection by appealing to Hobbes’s concurrentism.<sup>137</sup> Consider someone who presently desires posthumous fame, and who would acquire it by martyring himself tomorrow. The fulfilment of his desire would at no point contribute to his well-being: he will be dead and so cannot concurrently desire the fame, or experience any pleasure of satisfaction. It might therefore seem that the prospect of post-mortem fame fails to provide any present normative practical reason to secure its fulfilment. The same might seem to apply to any post-mortem passion, whether self-destructive (such as the desire for post-mortem martyrdom fame) or not (such as the desire for the post-mortem fame of

<sup>136</sup> L 26.21: 430.    <sup>137</sup> Meyers (2013).

having lived longer than even Hobbes). The same applies to the self-destructive desire to be dead: it is impossible for the desire to be concurrently satisfied.

Hobbes's concurrentism does not wholly dispense with the objection, however. First, securing conditions that would lead one to *anticipate* the future fulfilment of post-mortem or self-destructive passions might significantly contribute to one's well-being while still alive, by providing anticipatory pleasure.

Second, some self-destructive passions can be satisfied *before* death, and doing so might significantly enhance one's well-being while alive. Consider specifically *suicidal* passions, i.e., self-destructive desires and aversions that are conceptually linked to the desire to be dead or aversion to being alive (at some point).<sup>138</sup> Take the desire to be bleeding to death, or to be suffering a gruesome public killing (Murphy's example), or, most generally, to be about to die. Unlike the desire to be dead, one can have these suicidal desires at the moment of satisfaction. One can also be alive while in the grips of a suicidal aversion to being alive. Consider also non-suicidal self-destructive passions. Take someone who presently desires to exact revenge tomorrow, but who would be foreseeably put to death afterwards by his sovereign for doing so. If he will continue to desire revenge tomorrow, then exacting revenge might enhance his well-being.

It therefore seems quite possible for self-destructive passions to furnish or track normative reasons to engage in self-destructive actions. The real question is whether such reasons could, given Hobbes's strong prudentialism, ever be *decisive*, or whether instead they are always outweighed by other reasons. It is of course logically possible that the hastened-death scenario would maximize one's ongoing, lifetime well-being relative to other life scenarios.<sup>139</sup> Yet Hobbes's account of practical reasons turns not on what is logically possible or what will actually happen, but on what could be reasonably known or foreseen. Even if Hobbesian agents were maximizers of some kind – and there are grounds for doubt on this score<sup>140</sup> – we cannot interpret Hobbes's strong prudentialism in terms of a utility-maximizing principle; any such principle must be formulated in terms of what is reasonably knowable and foreseeable.

It is a rare life in which fulfilling self-destructive passions would actually maximize felicity. There are only two basic such scenarios. The first is where life afterwards, should death not be hastened, would be so wretched

<sup>138</sup> I adopt a more narrow definition of suicide than Brandt (1992: 315).      <sup>139</sup> Brandt (1992).

<sup>140</sup> Kavka (1986: 202–207).

that one would be, so to speak, better off dead. This would be a life inundated by immense misery, lacking even in the most basic “contentments.” The second scenario is where one would have died very shortly afterwards anyways, such that, even if death were not hastened a little by acting on self-destructive passions, one would have few occasions for having and fulfilling other future desires or experiencing future pleasures.

If one could reasonably know or foresee that life ahead is so wretched so as not to be worth living, or that death is imminent anyways, then one could have a decisive reason to desire and hasten one’s death. Hobbes was not particularly concerned about *suicidal* passions, given how rarely determinative he took them to be, psychologically; he was more concerned with non-suicidal self-destructive passions, which he thought frequently determinative. But reflecting on suicidal passions remains important because whether one has a decisive normative reason to have or act on a self-destructive passion in general turns out to mirror the question of whether one has a decisive reason to have and act on specifically suicidal passions. The issue is not whether one’s life will *actually* turn out to be wretched, but whether one has any reasonable hope of avoiding such ongoing, abject misery. If one does not, then committing suicide may be a reasonable course of action.

To be sure, Hobbes dismissed many cases of suicidal passion as manifesting the “Madnesse commonly called MELANCHOLY,” which consists in being subjected to “causelesse fears” by a violent or persisting “great *Dejection* of mind.”<sup>141</sup> This is precisely the “fit of madnesse” that had “seized” the “young Maidens” who had taken to hanging themselves in Miletus.<sup>142</sup> Such cases of irrational melancholy furnish the basis of the implicit argument, by the “philosopher” in Hobbes’s *A Dialogue Between a Philosopher and a Student, of the Common Laws of England* (1681), for the decriminalization of suicide. If anyone kills himself, Hobbes there suggested, “it is to be presumed that he is not *compos mentis*, but by some inward Torment or Apprehension of somewhat worse than Death, Distracted.”<sup>143</sup>

Hobbes did not, however, dismiss all cases of suicidal passion as manifestations of irrational melancholy.<sup>144</sup> “May not I rather dye, if I think fit?”

<sup>141</sup> L 8.18–20: 112.   <sup>142</sup> L 8.25: 116; LL 8: 117.

<sup>143</sup> D: 85. See Stoffell (1991), but also Chapter 5, note 122.

<sup>144</sup> Consider the Ethiopian kings who would commit suicide on the command of priests: their actions are “the strangest thing of all” and stem from “having their reason mastered by superstition.” This is not a case of melancholy, but perhaps another type of madness. B 2: 230, italics removed. Cf. HE 325–327.

Hobbes asked in *Questions*.<sup>145</sup> Even the cited passage from *Dialogue*, despite arguing for the *legal presumption* of insanity, appears to leave open the possibility that individuals might be tormented by something they have a good reason to believe is, and may actually be, “worse than Death.” Hobbes knew that suicidal passions sometimes strike because one can foresee only a wretched future. He may even have been aware of this from personal experience, if we are to trust the report of Gui Patin, the physician who attended to Hobbes at the end of his stay in Paris, and who found the philosopher “avec de telles douleurs qu’il avoit voulu se tuer.”<sup>146</sup> Yet in a rare discussion of such cases, Hobbes focussed on the fact that the foresight in question tends to be myopic:

the greatest [*primum*] of goods for each is his own preservation. For nature is arranged so that all desire their own good . . . By contrast, Death is the greatest of all Evils, especially when accompanied by torture; but so great can life’s pains [or grief, anxieties, melancholy, diseases: *ægritudines*<sup>147</sup>] be that, unless their quick end is foreseen, they lead Death to be numbered amongst the Goods.<sup>148</sup>

Although people may become suicidal because they cannot foresee a quick end to their miseries, it is nevertheless contrary to reason for them to (will to) end their life in order to avoid short-term miseries if their reasonably foreseeable *long-term* future would make those miseries worth enduring. Suicide remains unreasonable as long as some hope of relative felicity remains reasonable; suicide is reasonable only when there is no reasonable hope of escaping abject misery. But the conditions for reasonable hope are minimal: it consists in nothing more than a reasonable prospect of survival along with the minimal conditions of commodious living. This is why almost all lives are worth living, and why rational agents could each reasonably expect that theirs is one such life. Similarly, the hour of death is unknown to most, and few can reasonably know or foresee their own imminent death. If one cannot reasonably know or foresee that one’s life will be so wretched in the long run as to be not worth living, or that one’s own death is imminent, then self-destructive desires will not provide

<sup>145</sup> Q: 136/180.

<sup>146</sup> Patin (1846: 594). Patin himself, however, characterized Hobbes’s state as “mélancolique.”

<sup>147</sup> Although Cotton, *De Cive*’s translator, rendered Hobbes’s phrase *animi ægritudo* as “disease of the mind” (DCv 12.10), Hobbes was using the phrase as his Latin equivalent for what, in the corresponding passage in *Elements*, he called “paine,” “Discontent,” or “trouble of the minde” (EL 27.2–3). Hobbes did, by contrast, use *ægritudine* to translate *Leviathan*’s “Sickness” into Latin (L 38.15: 720; LL 38: 721).

<sup>148</sup> DH 11.6.

decisive practical reasons: acting on the assumption that one will be better off dead or that one would have died soon afterwards anyways is contrary to reason.

This brings us to the exceptions. Perhaps one is painfully and mortally ill and so can reasonably foresee both a miserable future and imminent death. Or perhaps one has committed a capital crime and can reasonably foresee an imminent future of torture and death at the sovereign's hands.<sup>149</sup> These cases render moot the likelihood that, in the long run, one's life will turn for the better: there will likely be no long run. In such circumstances, one may very well have a *decisive* reason to fulfil an intense self-destructive passion if it could be satisfied before one's hastened death. Or perhaps one has been forced, by command of the sovereign, to be the executioner of one's own parent: since one will have no prospect of undoing the past, one can reasonably know that one will never escape the abject misery of being a patricide: "a Son will rather die, then live infamous, and hated [or: in infamy and loathing (*infamis atque exosus*)]" for having committed such a deed.<sup>150</sup>

Hence we return to the problem with which we began: it looks as if the prudentialist interpretation, which grounds natural law in a principle of *felicity* rather than bare survival, cannot ground any completely universal and decisive reason to desire ongoing survival and avoid death. But this limitation, far from being an exegetical liability, turns out to be a merit: it reflects precisely Hobbes's view. There is no universally decisive normative reason to avoid death in Hobbes's ethics: Hobbes conceded that it may be reasonable to end a life not worth living, i.e., a life with no reasonable hope of escaping abject misery. There is a universally decisive reason to desire and to seek *self-preservation*, to be sure, but 'self-preservation' is not for Hobbes the synonym of 'survival' or antonym of 'death'. To return to the oft-quoted passage from *De Cive*, Hobbes asserted that it is not "*contra rectam rationem* for a man to expend all his efforts to preserve and defend his Body, and the Members thereof from death and sorrowes [*morte & doloribus*]."<sup>151</sup> 'Self-preservation' is the antonym of 'death-or-misery'; the counterpart to the desire for self-preservation is not the fear of death, but the fear of death-or-misery.<sup>152</sup>

The universally rational end that Hobbes posited as the ultimate basis for the laws of nature is therefore not (as is widely supposed<sup>153</sup>) bare

<sup>149</sup> L 21.17: 340.   <sup>150</sup> DCv 6.13.   <sup>151</sup> DCv 1.7. Cf. EL 14.6; L 13.3: 190.

<sup>152</sup> Cf. EL 19.2, which writes of "every mans being and wellbeing."

<sup>153</sup> Murphy (1995: 866). An important exception is Olsthoorn (2014).

survival, but self-preservation in the sense of the preservation of *a life worth living*. Natural law is grounded in “the security of a mans person, in his life, and in the means of so preserving life, as not to be weary of it.” Securing such a life is the universally rational “motive, and end for which this renouncing, and transferring or Right” to enter a commonwealth, as required by the second law of nature, “is introduced.”<sup>154</sup> This is why when “men that seek peace . . . lay down certaine Rights of Nature,” they must also “retaine some; as right to governe their owne bodies; enjoy aire, water, motion, waies to go from place to place; and all things else, without which a man cannot live, or not live well.”<sup>155</sup> It is also why the “finall Cause, End, or Designe of men” that grounds the law of nature’s prescription to join a commonwealth “is the foresight of their own preservation, and of a more contented life thereby; that is to say, of getting themselves out from that miserable condition of Warre.”<sup>156</sup> It is also why when Hobbes asserted that the sovereign is “obliged by the Law of Nature” to procure “the end, for which he was trusted with Sovereign Power, namely . . . *the safety of the people,*” he felt compelled to clarify that “by Safety here, is not meant a bare Preservation, but also all other Contentments of life, which every man by lawfull Industry, without danger, or hurt to the Common-wealth, shall acquire to himselfe.”<sup>157</sup>

The laws of nature do not presuppose, as countless commentators have assumed, that each has a decisive reason to avoid death at all costs. They presuppose, rather, that each has a decisive reason to desire self-preservation, not in the “bare” but in the wider sense of preserving a life of which one is not “weary,” a life lived minimally “well,” a life “more contented” than one condemned to “miserable” conditions until death – in short, a life worth living.<sup>158</sup> Natural law dictates leaving the state of nature because a life permanently condemned to it would be – and can be *known* to be – either a miserable life not worth living, or a short life, or both. This explains why what disposes humans to obey natural law is not simply the aversion to death. Rather, the “Passions that encline men to Peace, are [1] Feare of Death; [2] Desire of such things as are necessary to commodious living; and [3] a Hope by their Industry to obtain them.”<sup>159</sup> Chapter 13 of *Leviathan* makes the point succinctly and famously: the problem with “the

<sup>154</sup> L 14.8: 202.    <sup>155</sup> L 15.22: 234.    <sup>156</sup> L 17.1: 254.

<sup>157</sup> L 30.1: 520. Cf. EL 28.1. In L 11.1: 150, Hobbes argued that to consider “those qualities of man-kind, that concern their living together in Peace, and Unity” requires considering the nature of felicity.

<sup>158</sup> See Hobbes’s remarks about the “generall inclination of all mankind” to secure “the power and means to live well.” L 11.2: 150.

<sup>159</sup> L 13.14: 196.

life of man” condemned to the state of nature is that it would be (not merely short but) “solitary, poore, nasty, brutish, and short.”<sup>160</sup> Anyone who reduces the normative foundation of natural law to mere survival has failed to pause sufficiently to meditate on the first four words of Hobbes’s most famous of phrases.<sup>161</sup>

It is therefore this wider sense of a life worth living that Hobbes was invoking when in *Leviathan* he defined natural law as a rational precept forbidding each agent “that, which is destructive of his life” and prescribing that by which “it may be best preserved.” The wider sense of self-preservation and self-destruction explains why, in the Latin edition, Hobbes’s truncated translation speaks (not of what foreseeably destroys one’s life but) of what (foreseeably) tends to one’s *damnum*.<sup>162</sup> The normative foundation of both self-preservation and natural laws prescribing its means consists in the Prudential Precepts that prescribe desiring and securing one’s *felicity*.

The wider sense of self-preservation also explains the conditions under which the laws of nature articulate reasons for *action*. Although one ought under all circumstances to desire or be disposed to conform to the virtues articulated in natural law *in foro interno*, one ought not always act *in foro externo*. The laws of nature intrinsically dictate “Vertues,” “habits of the mind,” or “qualities that dispose men to peace.”<sup>163</sup> One does not have a decisive reason to *act* peacefully (first law of nature), *actually* to lay down one’s natural right to all things (second law), or *actually* to keep one’s covenants (third law), etc., if one could reasonably know or foresee that doing so would destroy one’s life – in the sense either of ending it or rendering it permanently odious. To act peacefully “in such time, and place, where no man els should do so, should but make himselfe a prey to others, and procure his own certain ruine, contrary to the ground of all Lawes of Nature, which tend to Natures preservation.”<sup>164</sup> This limit is embedded in natural law itself, which prescribes actions only *conditional* on it being safe to so act: it obliges “in effect then onely where there is Security.”<sup>165</sup> To fail to act peacefully when others would have replied with mortal aggression anyways is not to violate natural law at all. Hobbes’s position is therefore not that far removed from the Foole’s: the disagreement turns, first, on Hobbes’s assertion that one always has a decisive

<sup>160</sup> L 13.9: 192.   <sup>161</sup> Rutherford (2003) makes a similar point.   <sup>162</sup> L 14.3: 198; LL 14: 199.

<sup>163</sup> L 26.36: 442; 26.8: 418. Cf. DCv 3.29; Boonin-Vail (1994).

<sup>164</sup> L 15.36: 240. See DCv 3.29; 4.21.

<sup>165</sup> L 15.36: 240, marginal heading. The general precondition is security, not, as Kavka (1986: 344–347) has claimed, reciprocity. Lloyd (2009: 31–32).

reason to be *disposed* to fulfil covenants and, second, on his assertion that, under circumstances of security, even if violating a covenant would *actually* be advantageous but one cannot reasonably know or foresee this, then one still has a decisive reason to keep one's covenants.

To conclude: Hobbes was neither a subjectivist nor an instrumentalist about normative reasons. The conative-subjectivist interpretation would render Hobbes unable to hold, as he does hold, that we have a reason to take the relevant means to our normative ends; an instrumentalist interpretation would mistakenly have him say that we have a reason only to promote our *present* good, not our ongoing, overall good; and a cognitively-subjectivist interpretation would have him hold that natural laws do not furnish reasons for those who have misunderstood them. Hobbes was at least a prudentialist who thought that agents ought to care about their ongoing felicity, regardless of their current desires or beliefs. He was committed to the normativity of at least two types of objective foundational precepts, concerning the transmission of reasons from normative ends to relevant means, and concerning reasons grounded in one's ongoing felicity. Hobbes's prudentialism is presupposed by, and is necessary to explain, the normative status he attributed to self-preservation and the laws of nature, as well as the limits he placed on the normative imperative for sheer survival. The normativity of the Prudential Precepts is the basis for reasons of the good. The question therefore is: In what does an agent's good or felicity consist?

## *A Theory of the Good: Felicity by Anticipatory Pleasure*

One of the central assumptions of ancient Greek ethics is that human beings have a single supreme and ultimate good – *eudaimonia* or “well-being” – which includes and integrates all other final goods into an account of the good life.<sup>1</sup> As Cicero understood it, eudaimonist ethical theories also assume that the supreme good serves as the final end of every valuable action. This is why he began his survey of Greek ethics, *De Finibus Bonorum et Malorum*, by asking

what is the end [*finis*], which is final [*extremum*] and ultimate [*ultimum*], to which all our deliberations on living well and acting rightly should be directed? What does nature pursue as the supreme [*summum*] good to be sought, and what does she shun as the utmost evil [*extremum malorum*]?<sup>2</sup>

Cicero then proceeded to consider three rival eudaimonist theories of the good life: the Epicurean view, according to which *felicitas* – as *eudaimonia* was rendered into Latin – consists in a life of pleasure and freedom from pain; the Stoic view, according to which it consists in a life of virtue; and the Peripatetic view, according to which it consists in a life of virtue combined with bodily and external goods.

I shall argue Hobbes agreed with the Epicureans that *felicitas* consists in a life of pleasure and freedom from pain,<sup>3</sup> but departed from Epicurean hedonism in two significant ways. First, he rejected a key aspect of Epicurean eudaimonism: the assumption that felicity must be the ultimate *aim* of all valuable action, in the sense that it must be the intentional object of the mental state giving rise to action. This assumption had also been rejected by the Cyrenaic hedonists, who alone amongst the ancient philosophers had departed from eudaimonism, and who took the aim of all valuable action to be *particular* episodes of pleasure rather than felicity as

<sup>1</sup> Annas (1993).    <sup>2</sup> DF I.ii.

<sup>3</sup> For the sources of Epicurean influence on Hobbes, see Friedle (2012).

a whole. But Hobbes was no Cyrenaic either. He rejected the assumption that all actions are or should be aimed at any single final good – whether felicity *or* pleasure. For Hobbes, pleasure and felicity are typically the by-products of desiring and pursuing other aims. The Cyrenaics’ wholesale rejection of eudaimonism entailed denying that felicity – which they understood to consist in no more than the aggregation of individual episodes of pleasure (discounted by episodes of pain) – gives any coherent ordering of pleasures.<sup>4</sup> Hobbes, by contrast, was a eudaimonist who held that felicity is an overarching good giving coherence to a valuable life; he therefore agreed with the Epicureans that it is good to forgo lesser pleasures for greater future ones. He also rejected the Cyrenaic assumption that felicity simply measures the aggregate amount of pleasure (minus pain) over the course of one’s life: felicity concerns not just the *aggregate* amount of net pleasure, but its relative *distribution* or continuity over time. As he put it in *Leviathan*, the proper “object of mans desire, is not to enjoy once onely, and for one instant of time; but to assure for ever, the way of his future desire.”<sup>5</sup> Hobbes defined the felicitous life as a life of *ongoing* or – to quote *Elements* – “continual delight.”<sup>6</sup>

Second, Hobbes also rejected the assumption that felicity is a final end in the sense that attaining it would mark the *termination* of desire. Like the Stoics, the Epicureans held that felicity comprises an attainable and ongoing state of repose or tranquility (*ataraxia*) – which, according to the Epicureans, consists in the pleasure of being free from the pain of unsatisfied desires.<sup>7</sup> For Hobbes, by contrast, unsatisfied desires are not intrinsically painful, the absence of pain does not itself constitute a type of pleasure, and felicity is not a state of rest but is an ongoing motion or progression from pleasure to pleasure. It is not characterized by absence of desire. Hobbes was here echoing the Cyrenaics, even repeating the Cyrenaic criticism – which he would have encountered in Diogenes Laertius’s *Lives of Eminent Philosophers* – that tranquillity is the state of a corpse, not a life.<sup>8</sup> As he put it in *Leviathan*, “there is no such thing as perpetuall Tranquility of mind, while we live here; because Life it selfe is but Motion, and can never be without Desire, nor without Feare, no more than without Sense.”<sup>9</sup>

But Hobbes did not reject, as he has often been taken to have done,<sup>10</sup> the notion of a *summum bonum* or felicity as such. The relevant passages

<sup>4</sup> *Lives* I.2.8.87–88. See Mitsis (1988: 55); Annas (1993: 227–235). <sup>5</sup> L II.1: 150. <sup>6</sup> EL 9.7.

<sup>7</sup> Mitsis (1988); Annas (1993). <sup>8</sup> *Lives* I.2.8.89. Cf. Plato’s *Gorgias* 492e. <sup>9</sup> L 6.58: 96.

<sup>10</sup> Hampton (1992: 333); Schneewind (1998).

express instead his two key departures from Epicureanism: his denial that felicity is the utmost *aim* of every valuable action, and his denial that it is composed of *ataraxia* or the pain-free state of having no unsatisfied desires, i.e., the termination of desire. Hobbes was committed to the intrinsic value of pleasure and, in line with his eudaimonism about the good, to felicity as the overarching and ultimate value of a human life. Unlike the Cyrenaics, who held that the aim of every action is some individual episode of pleasure, and that one ought not to be concerned with one's overall felicity, Hobbes's prudentialism committed him to treating felicity as the ultimate ground for the value of one's actions and, indeed, as *a* proper "object of mans desire" in its own right.<sup>11</sup>

My agenda in this chapter is to defend this interpretation of Hobbes's theory of the good against rival ones that take him to have held either that there is nothing of intrinsic or of ultimate<sup>12</sup> value for agents – only ultimate bads – or that agents' ultimate good consists in the objective satisfaction of their desires. Hobbes did posit an ultimate and overarching good for a human life, which he conceived not as a final end or state to be realized but as an ongoing process of experiencing greater pleasures relative to pains. Felicity for Hobbes consists not primarily in the *fulfilment* of one's desires but in ongoing anticipatory pleasure engendered by the experience of repeated past success in satisfying one's desires. Whereas the Epicureans held that felicity consists in the tranquil state of *having satisfied* one's desires, and the Cyrenaics held that it consists in the somatic pleasures that arise while *satisfying* one's desires, Hobbes held that felicity consists primarily in the mental pleasures that arise from *anticipating* the satisfaction of one's desires.

#### 4.1 Four Distinct Questions about Goodness and 'Good'

To get a handle on Hobbes's theory of the good, we must carefully keep apart four questions: (1) In what does the property of goodness *consist*? (2) What *makes* something good (i.e., in virtue of what features does something have or acquire the property of goodness)? (3) What *causes* agents to see something as good and/or to call it "good"? And (4) what does the term 'good' *mean*? The first two questions are metaphysical and concern a substantive theory of the good, the third question is psychological, and

<sup>11</sup> L II.I: 150.

<sup>12</sup> Ultimate values are independent of anything else's value; intrinsic values are either ultimately or constitutively valuable (Korsgaard 1996a: chapter 9).

the fourth semantic. It is sometimes thought that these distinctions concerning the good, especially the distinction between metaphysical and semantic questions, are recent, twentieth-century inventions. But that thought is owed to intellectual amnesia. Even a cursory perusal of Cicero's *De Finibus* would dissipate it: the central issue in his comparison of the Stoic and Peripatetic doctrines concerns the "great controversy," brought to a head by Carneades, of whether "the dispute between these schools is substantial rather than verbal," i.e., whether the Stoics' theory of the good is substantively different from the Peripatetics', or whether the difference lies only in the meanings they assign to terms such as 'good'.<sup>13</sup> In his treatment of Epicureanism, Cicero had even highlighted the difference between the meaning of apparently normative terms in "everyday language" and the technical definitions given to them by moral philosophers.<sup>14</sup> That the four questions I have outlined are distinct does not imply they are unrelated. Because Hobbes considered 'good' to be one of the fundamental terms of his science of ethics, understanding his answer to the fourth, semantic question requires carefully distinguishing the term's customary, pre-scientific meaning as Hobbes understood it from the "apt" or *reforming* definition he proposed in light of his substantive, scientific account of goodness.<sup>15</sup> How Hobbes understood the term's customary meaning depends on his answers to the psychological question, but the reforming, scientific definition he proposed depends on his answers to the two metaphysical questions.

We should begin with a preliminary observation about the first question. Aristotle had noted that things are customarily called "good" in two senses: good unconditionally (*haplos*) or good for somebody.<sup>16</sup> Hobbes rejected the first usage as unscientific. Although he recognized that by custom people often use 'good' to mean what is good *simpliciter*, he insisted that the property of goodness is a *relational* property: a good thing is always good *for* someone in particular. In *Elements*, he flatly denied that there is "any such thing as ἀγαθὸν ἀπλῶς [*agathon haplos*] that is to say simply good," a point he illustrated with the dramatic example that "even the goodness which we attribute to God Almighty, is his goodnesse to us."<sup>17</sup> He reiterated in *Leviathan* that there is "nothing simply and absolutely" good,<sup>18</sup> and in *Questions* we find that even to speak of a good horse, in a scientifically informed way, is to speak of its goodness for some

<sup>13</sup> DF III.41; see IV.2; 21; 60.    <sup>14</sup> DF II.48–50.

<sup>15</sup> Scientific reasoning requires "apt imposing of Names." L 5.17: 72.    <sup>16</sup> NE 7.12.    <sup>17</sup> EL 7.3.

<sup>18</sup> L 6.7: 80.

individual(s).<sup>19</sup> *De Homine* clarifies that “since whatsoever is Good, is Good for someone or other,” nothing is *simpliciter Bonum*; nevertheless, there can be things that are a common good for many (*communiter Bonum*) or, indeed, a universal good for everyone (*omnibus bonum*).<sup>20</sup>

In denying that anything is “absolutely” good Hobbes was not claiming that nothing has *ultimate* value for agents; he was claiming that nothing has value apart from its value *for* agents. This assertion that goodness is relational does not furnish a full answer to our first question, about the nature of goodness, but it does yield enough to shed light on our second question, about what makes something good. The relational nature of goodness suggests that something is good in virtue of contributing to an *agent’s* ultimate good.

#### 4.2 The Meaning of ‘Good’: The Customary versus Reforming Sense

The discussion of the word ‘good’ in [chapter 7](#) of *Elements* proceeds in terms of pleasure: “Every man for his owne part calleth that which pleaseth and is delightful to himselfe GOOD, and that EVILL which displeaseth him.”<sup>21</sup> By contrast, the better-known, corresponding formulation in [chapter 6](#) of *Leviathan* is in terms of desire: “whatsoever is the object of any mans Appetite or Desire; that is it, which he for his part calleth *Good*: And the object of his Hate, and Aversion, *Evill*.”<sup>22</sup> Some commentators have suggested this difference reveals a shift in Hobbes’s substantive theory of value, from a hedonist theory in 1640 to a desire-fulfilment theory by 1651.<sup>23</sup> On the hedonist theory, agents’ well-being consists in pleasure (and absence of pain); on the desire-fulfilment theory, it consists in the objective satisfaction of desires (and non-frustration of aversions). The two theories come apart in cases in which desire is fulfilled without awareness: if I desire that my commonwealth’s troops succeed in battle and unbeknownst to me they do – perhaps I erroneously think they have been routed – then on the desire-fulfilment theory I am still better off, thanks to the objective satisfaction of my desire; but on the hedonist theory I am not better off in virtue of my commonwealth’s military success, since my desire is not *subjectively* satisfied and there is no resulting pleasure of satisfaction. The theories also come apart when one falsely believes one’s desires are

<sup>19</sup> Q: 146/192. <sup>20</sup> DH 11.4. See AW 30.26. <sup>21</sup> EL 7.3. <sup>22</sup> L 6.7: 80.

<sup>23</sup> McNeilly (1968: 120–121; 129–136; cf. 139, 145) argues that Hobbes’s characterization of felicity in [chapter 6](#) reflects a shift to the desire-fulfilment theory, but that his characterization in [chapter 11](#) harks back to the older, hedonist (and egoist) account. The desire-fulfilment theory is attributed to Hobbes by Gauthier (1979); Hampton (1986: 27–42, 53); Ripstein (1987: 124); Meyers (2013: 273).

fulfilled. Finally, on the desire-fulfilment theory, merely *having* a desire enhances one's well-being only if (and only in virtue of the fact that) one has a further, second-order desire to have the desire. By contrast, on the hedonist account, granting Hobbes's assertion that desire is always accompanied by anticipatory pleasure,<sup>24</sup> then merely having a desire – any desire – itself enhances well-being.<sup>25</sup> (However, if desire is not accompanied by hope, then it may also give rise to fear of failure and hence the anticipatory pleasure may be outweighed by the anticipatory pain of despair. Or if, as the Epicureans thought, unsatisfied desire is intrinsically accompanied by the pain of non-satisfaction, then having desires will damage, rather than promote, one's well-being.)

The alleged shift to a desire-fulfilment theory is presumably supposed to have begun during the composition of the first edition of *De Cive*, published in 1642, for it is here that we first find Hobbes characterizing 'good' in terms of appetite: "*Good* and *Evill* are names imposed upon things to signifie appetite for, or aversion to what are thus named."<sup>26</sup> We subsequently find kindred formulations in the 1645 treatise *Of Liberty and Necessity*, where Hobbes wrote that "to love a thing and to think it good are all one,"<sup>27</sup> and in *De Homine*, published in 1658, which reiterates that "The common name for all things that are desired, insofar as they are desired, is *Good*; & for things we eschew, *Evil*."<sup>28</sup> This interpretation thus involves two claims: first, that the passages in question record a shift in Hobbes's view; second, that the passages answer to our second question, about what makes things good.

Both elements are mistaken. The historical-shift hypothesis is belied by the fact that the two supposedly divergent formulations coexist throughout Hobbes's career. In *Anti-White*, the manuscript dated to 1642–1643, Hobbes's formulation appears to realign with *Elements*: "Everyone indeed calls good that which pleases the one pleased, that is, one's own good [*bonum sibi*] what pleases one, and the commonwealth's good what pleases the commonwealth."<sup>29</sup> And again: "everyone calls good what is pleasurable to himself [*sibi iucundum*], & likewise calls 'good man' simply, whomever he approves."<sup>30</sup> Moreover, we find an appeal to *both* desire and to pleasure in *De Cive*: "Such is the nature of man, that every one calls that *good* which he desires, and *evill* which he eschewes . . . For we all measure [*æstimamus*] *Good* & *evill* by our own delight & pain [*molestiâ*]."<sup>31</sup> Nor does the hypothesis that 1642–1643 was a transitional period, after which Hobbes

<sup>24</sup> L 6.11: 82.    <sup>25</sup> For a contemporary discussion, see (Brandt 1992).    <sup>26</sup> DCv 3.31.

<sup>27</sup> LN §33.    <sup>28</sup> DH 11.4.    <sup>29</sup> AW 31.1.    <sup>30</sup> AW 38.4.    <sup>31</sup> DCv 14.17.

settled on the desire-fulfilment theory first broached in *De Cive*, provide any refuge. For we find Hobbes alternating between the language of desire and pleasure once again in 1645, in *Of Liberty and Necessity*: “what is it to say an action is good but to say it is as I would wish, or as another would have it, or according to the will of the state, that is to say, according to the law? Does my Lord think that no action can please me or him or the commonwealth that should proceed from necessity?”<sup>32</sup> Indeed, the purported transitional period would have to extend all the way to 1656, since we find the language of pleasure deployed once again, five years after *Leviathan*, in *Questions*: “*Good* is relative to those that are pleased with it.”<sup>33</sup>

Instead of the hypothesis of a historical shift, the evidence across Hobbes’s texts therefore supports the unsurprising conclusion that he did not have the difference between a hedonist and desire-fulfilment theory in view. But that conclusion is consistent with at least three interpretive hypotheses: that Hobbes was muddled and equivocated between the hedonist and desire-fulfilment theories; that Hobbes was committed to a desire-fulfilment theory, and his talk of pleasure can be explained in its terms; or that Hobbes was essentially committed to a hedonist theory, and his talk of appetite can be explained in its terms. We have considerable evidence for the third hypothesis.

The evidence appears in light of the fact that the passages about the term ‘good’ quoted above do not answer to our second question, about what makes something good, but to our third question, about what causes individuals to *call* something good. Thus Hobbes wrote in *Elements* that “Every man for his owne part calleth that which pleaseth and is delightful to himselfe GOOD,” and in *Leviathan* that “whatsoever is the object of any mans Appetite or Desire; that is it, which he for his part calleth *Good*.” Once we realize Hobbes was engaged in psychological explanation here, his alternation between the language of pleasure and desire becomes perfectly understandable: it does not indicate any equivocation or shift, but rather the substantive equivalence of the alternative formulations. For on Hobbes’s account, desire and pleasure are two aspects of a single passion: desire is the conative aspect, i.e., the motion or *conatus* towards an intentional object that propels the agent to act, while pleasure is the cognitive or representational aspect, i.e., the “appearance” of the intentional object in which the object is represented as good. Desire is always accompanied by pleasure – both correspond to the same internal bodily motions – which is why Hobbes could write in *Elements* that terms such as ‘pleasure’ or

<sup>32</sup> LN §14.    <sup>33</sup> Q: 145/192.

‘delight’ and ‘desire’ or ‘appetite’ are but “diverse names for diverse Considerations of the same thinge.” Desire is “the Indeavour, or internall beginning of animall motion” towards some intentional object, and it is always desire for an object insofar as “the object delighteth.”<sup>34</sup>

In other words, Hobbes was a psychological egoist and hedonist of sorts, according to whom (a) all action is prompted by desire, (b) all desire is accompanied by pleasure, and (c) whenever we desire anything, we desire it only if we represent it as something pleasant to ourselves. As he put it in *De Cive*, “whatsoever a man would [*voluerit*], it therefore seems good to him because he wills it,” and “whatsoever seemes Good, is pleasant [*iucundum*], and pertains either to the senses, or the mind.”<sup>35</sup> Hobbes’s language might even seem to commit him, as many of his readers have thought, to a stronger form of psychological egoism and hedonism – the kind advanced by the Cyrenaics – according to which the *object* of every desire, and hence the aim of every action, is one’s own pleasure.<sup>36</sup> This statement in *Leviathan* is representative: “the proper object of every mans Will, is some Good to himselfe.”<sup>37</sup> But Hobbes in fact asserted that we can and do desire things other than our own pleasure: for example, benevolence is “*Desire* of good to another” (which rules out strong psychological egoism), and curiosity is a “*Desire* to know why, and how,” not a desire for pleasure (which rules out strong psychological hedonism). Despite the occasionally strong formulation, Hobbes did not intend to endorse psychological egoism and hedonism in the strong sense. His discussion of curiosity is especially telling because it echoes one of Cicero’s central arguments in *De Finibus* against strong psychological hedonism (which Cicero took the Cyrenaics and Epicurus to have defended). That argument appeals to the desire for knowledge: intellectual pursuits manifest an “innate love of learning and knowledge,” Cicero argued, and cannot be reduced to pursuits “for the sake of mental pleasure.” Rather, they are pursued for their own sake “precisely because the mind takes delight in them in the absence of any utility.”<sup>38</sup> Hobbes’s weak psychological hedonism evades the force of Cicero’s objection, and echoes Cicero’s point about the role of pleasure in explaining our non-hedonic desires. The point Hobbes was making in his apparently stronger formulations is that insofar as we desire things such as the good of another or knowledge, we do so only as a result of the fact that we see these objects as pleasurable – because we take, for example, “delight in the . . . generation of Knowledge.”<sup>39</sup> Although our own pleasure is not

<sup>34</sup> EL 7.2.    <sup>35</sup> DCV I.10; I.2.    <sup>36</sup> Nagel (1959).    <sup>37</sup> L 25.2: 398.    <sup>38</sup> DF V.48–50.

<sup>39</sup> L 6.22: 84; 6.35: 86.

the *object* of every desire, on Hobbes's mechanistic psychology the fact that an intentional object appears pleasurable to us is a necessary part of the *causal* process by which our desire for it arises. We do not necessarily desire the *pleasantness* of acquiring knowledge, but its pleasing aspect plays a causal role in explaining our desire for it.<sup>40</sup>

The result is that people are disposed to *call* the objects of their desire good, because in desiring them they not only experience pleasure but also conceive them as pleasurable and hence as good for themselves. This is why Hobbes could alternate between the language of pleasure and desire. The passages in question about 'good' articulate Hobbes's constant view about the *customary meaning* of the term, not a substantive, allegedly shifting, *theory* of the good: part of what it means to call something good in common discourse is to signify one's desire for it or, concomitantly, to signify one's conception of it as pleasant.

We can confirm the fact that Hobbes was answering our third, psychological question in these passages by more precisely attending to his distinction between the customary, prescientific sense of 'good' and his own scientific, reforming sense – a distinction grounded in Hobbes's rejection of the Aristotelian method as he understood it. That method consists in beginning with the common opinions expressed in the customary use of words, and then sifting through them dialectically to resolve contradictions and hence to arrive at purportedly warranted, scientific belief. Hobbes rejected this method as too beholden to custom: the failure of moral and political philosophy prior to his own, Hobbes asserted in *Elements*, stems from the fact that Aristotle and his successors "take for principles," and hence as the basis for the foundational definitions of science, "those opinions which are already vulgarly received, whether true, or false, being for the most part false."<sup>41</sup> This already suggests – although Hobbes did not say so explicitly in *Elements* – that the true philosopher must leave aside the common opinions reflected in customary word use.<sup>42</sup> He must instead directly examine the *conceptions* that words signify and reason with words whose signification he has reformed in light of his conceptual analysis. Hobbes explicitly drew this conclusion in *Leviathan*, declaring it necessary for the scientist "to examine the Definitions of former Authors; and either to correct them, where they

<sup>40</sup> See Gert (1965, 1967); Hampton (1986: 17–24); Boonin-Vail (1994: 42–47). Unlike Gert's "tautological" egoism, the weak egoism/hedonism combination has substantive behavioural consequences: we do not desire things we cannot see as pleasant. For criticism of Gert, see Olsthoorn (2014).

<sup>41</sup> EL 13.3. <sup>42</sup> As Cicero had suggested in DF IV.4.

are negligently set down; or to make them himselfe.”<sup>43</sup> Customary definitions do not provide foundations for science; their treatment is necessary only as a preliminary step for diagnosing where received opinion has gone wrong, via a scientific analysis of the illusions to which the mind is prone. This is partly what motivated Hobbes to hail Plato, rather than Aristotle, as “the best Philosopher of the Greeks.”<sup>44</sup> Whereas philosophers inspired by Aristotle have “reasoned only from the names of Things,” those inspired by Plato “founded their Doctrine upon the Conceptions and Ideas of things.”<sup>45</sup>

Hobbes’s distinction between the customary and reforming or scientific senses of ‘good’ is grounded in a further distinction, namely, between two types of pleasure that can cause people to call things good: pleasures of satisfaction or *fruition*, which arise from perceiving something *presently* satisfying one’s desire, and pleasures of imagination or *anticipation*, which arise from imagining one’s desire being satisfied – paradigmatically in the future. If merely *having* a desire involves anticipatory pleasure, then presently perceiving the object of one’s desire being satisfied yields a pleasure of fruition: if “the beginning of animall motion toward something which pleaseth us” is “Appetite” and hence involves anticipatory pleasure, then “the END of that motion” is “the attayning thereof,” and “when wee attaine that end the delight we have thereby is called FRUITION.”<sup>46</sup> It is true that when Hobbes came explicitly to distinguish between types of pleasure, his classification was in terms of pleasures of *sense* and *mind*, rather than pleasures of fruition and anticipation. When in *Elements* he declared “There are two sorts of pleasures,” he contrasted the “SENSUAL” kind, such as the pleasure of sex or eating, which “seemeth to affect the Corporiall organ of sense,” to “the Delight of the minde” or “JOY,” which “sort of delight is not particular to any part of the body.”<sup>47</sup> But from early on Hobbes equated *pleasures of sense with those of satisfaction, and pleasures of mind with those of anticipation*. Pleasures of sense involve “Conception of the present,” while pleasures of the mind “consist in Conception of the Future, that is to say in Conception of power past, and the Acte to come.”<sup>48</sup> Similarly, in *Leviathan* Hobbes defined “*Pleasures of Sense*” as those arising “from the sense of an object Present” to one’s sensory organs, while he defined “JOY” or the “*Pleasures of the mind*” as those arising from the “Expectation, that proceeds from foresight of the End, or Consequence of things.”<sup>49</sup>

<sup>43</sup> L 4.13: 56. See L 34.1: 610; DC 2.4; 6.15. <sup>44</sup> L 46.10: 1058.

<sup>45</sup> D 94. Strauss (1963: 139–141).

<sup>46</sup> EL 7.5.

<sup>47</sup> EL 7.9.

<sup>48</sup> EL 8.2–3.

<sup>49</sup> L 6.12: 84.

Hobbes's distinction between sensory pleasures and mental pleasures is rooted in ancient Greek ethics, and we can best appreciate the significance of his having collapsed it into the distinction between pleasures of satisfaction and anticipation by highlighting what differentiates Hobbes's account from his hedonist predecessors. Epicurus made two, cross-cutting distinctions between pleasures: he contrasted sensual or *somatic* pleasures to mental or *psychic* pleasures on the one hand and *kinetic* pleasures to *katastematic* ones on the other. Kinetic pleasures are pleasures of motion and consist in presently *satisfying* one's desire. Epicurus held that some kinetic pleasures are somatic (*euphronysé*), as when some object is presently titillating the relevant sensory organ, while other kinetic pleasures are psychic (*chara*), as when one reflects on kinetic pleasures of the body. Katastematic pleasures, by contrast, consist in the freedom from pain that comes either from *already having satisfied* or having *relinquished* one's desires or aversions. This type of pleasure can either be somatic (*aponia*), as when one has satisfied one's thirst, or it can be psychic (*ataraxia*), as when one has relinquished the fear of death. Epicurus taught that felicity is composed of katastematic, not kinetic, pleasure.<sup>50</sup>

Cicero translated 'katastematic' into Latin as 'static' (*stante*) because he took such pleasures to consist in the absence of motion; he also interpreted Epicurus in such a way as to equate kinetic with sensual pleasures, and static with mental pleasures. As a result, on the Ciceronian interpretation of Epicureanism – which Hobbes took for granted – felicity wholly consists in a static condition of repose or tranquillity of the mind.<sup>51</sup> Hobbes opposed this Epicurean view in part because, like the Cyrenaics and Cicero, he rejected the notion of static pleasures: he took pleasure to arise from motion, such that pleasure consists neither in having satisfied or relinquished a desire nor in the absence of pain. He understood all sensory pleasures to be "kinetic" pleasures that persist only so long as one is presently *satisfying* one's desire. This is what a pleasure of fruition is for Hobbes. A Hobbesian pleasure of mind, by contrast, does not wholly map onto either katastematic or kinetic pleasure: it is like kinetic pleasure in that it arises from motion, but it is unlike kinetic pleasure in that it arises not from presently satisfying one's desire, but from *imagining* one's desire being satisfied. Just as imaginations are parasitic on sensory perceptions, pleasures of mind are parasitic on pleasures of sense: whether "foresight of the End, or Consequence of things" yields a mental pleasure or mental pain depends on "whether those things in the Sense Please or Displease."<sup>52</sup>

<sup>50</sup> *Lives* II.10.136. Wolfsdorf (2013: 147–149).    <sup>51</sup> DF II.10. Annas (1993: 336).    <sup>52</sup> L 6.12: 84.

Because Hobbes took the process of *satisfying* one's desire to be constituted by the titillation of senses, pleasures of mind ultimately arise from imagining a satisfying process that yields sensory pleasures.

It might be thought that cases of satisfying a desire to *imagine* something run counter to Hobbes's equation of pleasures of sense with fruition and pleasures of mind with anticipation. For as Hobbes was aware, mental pleasures sometimes arise from "feigning or supposing" things and their imaginary consequences, such as joys "nourished by the Histories, or Fictions of Gallant Persons," or the pleasure of "imagining, or dreaming of the death of him, from whose life he expecteth nothing but dammage." Hobbes deemed the phenomenon ubiquitous: "to be pleased in the fiction of that, which would please a man if it were reall, is a Passion so adhaerent to the Nature both of a man, and every other living creature."<sup>53</sup> Thus the anticipation that produces joy might comprise either a hope of satisfaction, or a purely fictional "Expectation of consequences" (or, indeed, a memory of past consequences).<sup>54</sup> But when a poet playfully *desires* to imagine himself king without desiring really to become king, and takes pleasure in imagining himself king – or when a man is "delighted in the Imagination onely, of being possessed of another mans goods, servants, or wife, without any intention to take them from him by force, or fraud"<sup>55</sup> – then satisfying the playful desire would appear to produce a pleasure of mind that is, at the same time, a pleasure of satisfaction – in defiance of Hobbes's equation of pleasures of satisfaction with pleasures of sense.

Hobbes would have denied, however, that such cases involve a pleasure of *satisfaction*: to imagine oneself king may be pleasant, but the pleasure is purely of expectation. One experiences pleasure here not in *satisfying* one's desire but in virtue of imagining and hence *anticipating* something pleasant. Even while satisfying a desire *really* to become king, any accompanying pleasures of satisfaction would arise from the more specific desires for sensory experiences that make up the complex desire to become king (such as the desire to bear a crown, to see oneself hailed, etc.) Otherwise, Hobbes suggested the accompanying pleasures would not be of fruition – which are always pleasures of sense – but pleasures of mind: they would not arise from

<sup>53</sup> L 6.12: 84; 6.41: 88; 27.1: 452.

<sup>54</sup> Although mental pleasures paradigmatically concern the future, they may also concern the *past* or, indeed, mere fictions, because for Hobbes both past and future are fictions: "expectation of the future . . . is the same as memory of the past." Hobbes nevertheless prioritized the future: "What is in the memory, does not please as something past, except insofar as the past events are signs of similar pleasures expected in the future." AW 30.23; 38.5. Cf. EL 8.2.

<sup>55</sup> L 27.1: 452.

satisfying the desire to become king per se but from contemplating one's power to become king and the consequently hopeful anticipation of being able to satisfy further desires.

Hobbes did, however, acknowledge a somewhat hybrid case between pleasures of sense and pleasures of imagination – *somewhat* hybrid because it is ultimately a pleasure of imagination, but caused by present sensory perception. The hybrid case touches on a potential ambiguity behind describing a pleasing object as “present” or presently pleasing. On the one hand, a pleasing object may be “present” either in the temporal sense that it is presently satisfying one's desire, or in the spatial sense that it is present to the senses. These two senses can come apart: a parched desert traveller who had given up hope may suddenly *see* an oasis nearby, without yet *drinking* the water and quenching her thirst. In such a case the water she desires is an object of sensory perception, but her desire for it is not yet being satisfied since it is present to her *visual* rather than gustatory sensory organ. Although Hobbes characterized the intentional objects of desire as bodies, in reality the objects of desire are for him always bodies *under some particular description* – such as the traveller's desire for water *as* being drunk by her. Thus while the *water* is an object of sensory perception and hence is present to the traveller in the spatial sense, her desire for water-as-being-drunk-by-her is presently not yet being satisfied.<sup>56</sup>

On the other hand, something may be presently pleasing either in the sense that one's desire is being presently satisfied (yielding a pleasure of sense) or in the sense that one presently anticipates its future satisfaction (yielding a pleasure of mind). But the case in which the traveller sees the water she desires to drink is a case in which she experiences an anticipatory pleasure of mind in part *thanks to* her sensory perception of the water: seeing the water both nourishes the hope of satisfying her desire and intensifies her desire for water by making her representation of it more vivid. (Sensory perception is more vivid because imagination “is nothing but *deceiving sense*.”<sup>57</sup>) Seeing a physical object may even give birth to

<sup>56</sup> Although Hobbes strictly equated the object of mental states with *things* – bodies – he also loosely characterized intentional objects as events or states of affairs: “Sometimes a man desires to know the event [i.e., outcome] of an action; and then he thinketh of some like action past, and the events thereof one after another.” L 3.7: 42. “Action” here means event, and “event” means a state of affairs. Hobbes could speak in this way because events are “accidents” of bodies and states of affairs are composed of bodies in a given set of relations. To say that a man “thinketh” of some action is, strictly speaking, to say he thinks of a body-which-is-acting. The intentional object of passion is *always* implicitly a body conceived of *as* involved in some event or composing some state of affairs: to say one desires food (L 6.2: 78) is implicitly to say one desires to possess, taste, or eat the food. Cf. Kavka (1986: 295–296); Parfit (2011: vol. 1, 43–44).

<sup>57</sup> L 2.2: 26.

entirely new desires: the wayfarer wandering through a Persian orchard initially desires only to take in its sights, but in stumbling upon the pomegranate tree is overtaken by the desire to taste its unfamiliar scarlet fruit. Strictly speaking, the pleasure involved in such cases is a mental pleasure, but it is distinct from other cases of anticipatory pleasure in that it is partly grounded in sensory perception. One takes pleasure from simultaneously perceiving an object and imagining it in some as-yet unrealized state or event. As Hobbes put it in *Anti-White*, “delight and pain consist sometimes in sense, sometimes in imagination,” but also “sometimes in sense and imagination simultaneously.”<sup>58</sup> This hybrid type of anticipatory pleasure is significant for Hobbes because he took the desires associated with it – desires in part grounded in present sensory perception – to be more powerful motivators in general than desires associated with pure mental pleasure. This is precisely why, as he put it in *De Cive*, “men cannot put off this same irrationall appetite, whereby they greedily prefer the present good [*bona presentia*] before the future” – *presentia* either in the sense of presently satisfying one’s desire or of being present to the senses.<sup>59</sup>

It is this complex set of distinctions – (1) between (a) sensory pleasures arising from present satisfaction and (b) mental pleasures concerning future satisfaction, and (2) between (a) hybrid pleasures of anticipation and (b) pure mental pleasures of anticipation – that undergirds Hobbes’s distinction between the customary and scientific senses of ‘good’. This distinction quietly surfaces for the first time in the seventeenth chapter of *Elements*:

Every man by naturall passion, calleth that good which pleaseth him [1] for the present, or [2] so farre forth as he can foresee; and in like manner that which displeaseth him Evill. And therefore he that foreseeth the whole way to his preservation . . . must also call it good, and the Contrary, Evill. And this is that good, and evill, which not [1] every man in passion calleth soe, but [2] all men by reason.<sup>60</sup>

Hobbes was here making both an empirical claim about the customary use of the term ‘good’ and a substantive claim about goodness. On the one hand, the paragraph’s opening – which describes what people do “by naturall passion” – makes a claim about human psychology and language use: Hobbes was claiming that human beings call good either (1) what pleases them “for the present,” or (2) what pleases them about what they

<sup>58</sup> AW 30.23.    <sup>59</sup> DCv 3.32. Cf. LN §23.    <sup>60</sup> EL 17.14.

can anticipate, over the course of time. This is a psychological claim about what humans do, not a metaphysical claim about what the good consists in or what makes things good. On the other hand, Hobbes indicated that the second sense of ‘good’ has a privileged status: to call “good” what is pleasant for the *present* is to use the term merely “in passion,” but to call “good” what concerns “the whole way” one can foresee into the future is to speak “by reason.” As Hobbes noted in the immediately following sentence, natural laws are good in this latter, rational or scientific sense of the term: “the fulfilling of all these Lawes is good in reason.” To use the term ‘good’ merely “in passion” is to use it in its pre-scientific, customary sense, but to use it “by reason” is to use it according to the scientific sense by which Hobbes was proposing to fix the meaning of ‘good’. This scientific usage reflects Hobbes’s substantive theory of what *makes* things good.

The distinction between the two senses of ‘good’ resurfaces in *De Cive* in a more political context. Because whenever people desire something they conceive it as good and pleasant,<sup>61</sup> it is customary for them to call the objects of their desires “good”: “Such is the nature of man, that every one calls that *good* which he desires, and *evill* which he eschewes,” so that in practice “*Good* and *Evill* are names imposed upon things to signifie appetite for, or aversion to what are thus named.”<sup>62</sup> The problem with this customary use of ‘good’ – the genesis of which Hobbes’s psychological account is meant to explain – is that it cannot be the basis for any stable *agreement* about what to call good. Because “the appetites of men are diverse,” people are prompted to disagree about what to call good, so that “what this man *commend*s [Laudat], (that is to say, calls *Good*) the other *undervalues* [vituperat], as being *Evil*; Nay, very often the same man at diverse times, *praises*, and *dispraises* [laudat & culpat] the same thing.” The customary sense of ‘good’ therefore cannot serve either as the basis for a *science* of the good, or as the basis for peace. So long as individuals use ‘good’ and ‘evil’ in this customary way – to signify whatever appetites and aversions they happen to have at the time – they will face “discord, and strife” and hence find themselves in “the state of War.”<sup>63</sup>

There is, however, a rational basis for a solution to the problem of war and for establishing a reformed, scientific meaning for ‘good’. That basis lies once again in distinguishing the two types of pleasure that might prompt people to call something good: “we all measure *Good & evill* by our own delight & pain (which is either present, or expected).”<sup>64</sup> Those trapped in a state of nature are unable to agree on any common object of

<sup>61</sup> DCv 1.10; 1.2.    <sup>62</sup> DCv 14.17; 3.31.    <sup>63</sup> DCv 3.31.    <sup>64</sup> DCv 14.17.

*present* sensory pleasure or, indeed, of hybrid anticipatory pleasure from objects they can presently *perceive* via the senses, because, as long as they are in that state, “by reason of the diversity of the present appetites, they mete *good & evill* by diverse measures.” But it is possible to agree on an imagined common future object of *anticipatory* pleasure, since “All men easily acknowledge this state [which is a state of war], as long as they are in it, to be evill; & by consequence that *peace is good*.” The anticipatory pleasure raised in everyone’s mind by the prospect of peace and its pleasures is a potential basis for agreeing about what to call good. Calling peace and its relevant means “good” is therefore not only itself a potential basis for actually securing peace, it also corresponds to the term’s scientific meaning: “They therefore who could not agree concerning a present, doe agree concerning a future good; which indeed is a work of reason; for *things present* are perceived by the senses, *things future* by nothing except Reason.” Reason both engenders anticipatory pleasure concerning a future peace and itself declares “peace” and all the relevant “means to peace [to] be good.” Calling peace good therefore aligns with the reforming definition proposed by Hobbes’s philosophy. In this respect, Hobbes sharply contrasted his own philosophy to those of his predecessors, who had failed to “observe the goodness of actions to consist in this, that it was in order to Peace,” and who had consequently “built a *morall Philosophy* wholly estranged from the *morall law*, and unconstant to it self.”<sup>65</sup>

If in *Elements* and *De Cive* Hobbes’s distinction between the customary and scientific senses of ‘good’ rises to the surface only furtively, in *Leviathan* it finally lands and sets foot confidently ashore. Hobbes there reiterated his argument that, on the one hand, “Morall Philosophy is nothing else but the Science of what is *Good*, and *Evill*” and that, on the other hand, speakers customarily use the terms ‘good’ and ‘evil’ to signify whatever appetites and aversions they happen to have at the moment – which leads to semantic inconstancy, disagreement, and consequently war.<sup>66</sup> But he now also laid out with perfect clarity, early in *Leviathan*, the implications for any scientific undertaking of these “kinds of Speeches,” which “signifie the Appetites, Aversions, and Passions of mans mind”: “The names of such things as affect us, that is, which please, and displease us, because all men be not alike affected with the same thing, nor the same man at all times, are in the common discourses of men, of *inconstant* signification” and, as such, “can never be true grounds of any ratiocination.” Because words that, “besides the signification of what we

<sup>65</sup> DCv 3,31–32.      <sup>66</sup> L 15,40: 242.

imagine of their nature, have a signification also of the nature, disposition, and interest of the speaker” are disposed to inconstant signification, “therefore in reasoning, a man must take heed of [such] words” and begin by first firmly *settling* their scientific meaning.<sup>67</sup> Hence the need for reforming, scientific definitions.

Hobbes was proposing to settle or fix the meaning of ‘good’ such that it names not whatever people happen to desire or find pleasant at any given point in time – as the ancient Greek philosophers had supposedly done, whose “Morall Philosophy is but a description of their own Passions” and who “make the Rules of *Good*, and *Bad*, by their own *Liking*, and *Disliking*”<sup>68</sup> – but what is good for them in reason, their true, long-term good. Hobbes was proposing a reforming definition for ‘good’ firmly grounded in his substantive theory of what goodness consists in and of what makes it the case that something has that property.

### 4.3 What Makes a Life Good: Anticipatory Pleasure

Hobbes explicitly contrasted agents’ apparent good to their true good for the first time in *Anti-White*, noting that whether an action contemplated in deliberation is good for them overall depends on whether its total, long-run consequences involve a greater amount of pleasure relative to pain:

the things that please are connected with those that pain in a series so long that we are not able to foresee at one glance all the way to the end of the chain, & the connection between those that delight and pain, i.e., the good & evil, is so tight that they are taken up or relinquished together. If in this series there is more good than evil, the whole is good, so it is *better* to take up the whole, and worse to relinquish; if by contrast there is more evil than good, the whole is evil & it is worse to take up, better to relinquish.<sup>69</sup>

Hobbes followed up this evaluative point – which echoes Epicurus<sup>70</sup> – with the observation that as a matter of fact individuals will be disposed to undertake actions that merely *seem* to them to promise greater net pleasure. Because of limited foresight, the apparent goodness of actions will not

<sup>67</sup> L 4.23–24: 62. Philosophers must begin “their Ratiocination from Definitions; that is, from settled significations of their words.” L 5.8: 70; 5.4: 66.

<sup>68</sup> L 46.11: 1058–1060. Cf. EL 27.13. <sup>69</sup> AW 30.25.

<sup>70</sup> Hobbes would have encountered Epicurus’s so-called “hedonistic calculus” passage in *Lives* II.10.129. Cf. *Epicurus*, pp. 87–89. Some scholars deny the passage commits Epicurus to a pleasure-maximizing calculus, on the grounds that katastematic pleasures are neither quantifiable nor ordinarily rankable (Annas 1993: 85–87, 334–335).

always coincide with individuals' true good, i.e., with the action that will produce greater net pleasure for them in the long run:

We take up & relinquish according to the distance we foresee along some delimited [*determinatam*] expanse, in which there may be more good than evil, but in the part not foreseen there may be so much evil that in the whole there is more evil than good, because it is necessary to take up the whole, on account of the connection we take up not what is truly good, but what has appeared good, and this is what is called *apparent good*, which is sometimes evil and displeasing [i.e., overall].<sup>71</sup>

This distinction between *bonum verum* and *bonum apparens* appears again in *Leviathan* and *De Homine*,<sup>72</sup> but *Anti-White* remains the text in which Hobbes provided his clearest and most detailed exposition on the true good and on felicity. If Hobbes's discussion of *bonum verum* concerns the instrumental value of the actions contemplated in deliberation, felicity is what is ultimately valuable in a human life. There are two intrinsic constituents of felicity on Hobbes's account. The first reflects Hobbes's *hedonism*: the value of a life is intrinsically constituted by pleasure (and absence of pain): "felicity consists in living a life with pleasure, i.e., maximum delight [*jucunditate maxima*]." <sup>73</sup> Like the Cyrenaics, Hobbes took every experience of pleasure to be intrinsically good for agents – a constituent of their felicity – and whatever promotes their pleasure to be instrumentally good for them. The second intrinsic feature of felicity reflects Hobbes's *eudaimonism*. Hobbes's eudaimonism does not merely involve his agreement with the Epicureans that future pleasures count as much as present pleasures. More significantly, Hobbes conceived of felicity as an overarching good that integrates the value of discrete experiences of pleasure into a coherent whole. This is why he rejected the Cyrenaic assumption that felicity is a mere aggregation of pleasures. Hobbes assumed that the relative distribution of pleasures over time is an independent and second intrinsically valuable feature of a felicitous life: a life in which all pleasures are experienced in a single moment is less felicitous and therefore less valuable than one in which an equal amount of aggregate pleasure is distributed more evenly and hence continuously experienced throughout one's life. To be sure, Hobbes did not have a metric for adjudicating trade-offs between the aggregative versus temporally distributive dimensions of felicity. But he posited the ongoing or "continual" experience of pleasure as an intrinsic dimension of felicity in its own right. Felicity consists in pleasure "that is not brief" and, indeed, "lies not in the

<sup>71</sup> AW 30.25.    <sup>72</sup> L 6.57; 94; DH 11.5; 12.1.    <sup>73</sup> AW 38.8.

goal but in the path itself.”<sup>74</sup> Felicity is an integrative value: it consists in (a) pleasure (and absence of pain) (b) experienced on an ongoing basis continually over time.

This second, integrative dimension of felicity frames the whole of Hobbes’s inquiry in *Anti-White* and motivates his focus on anticipatory pleasures. Even if one grants that “felicity consists in living a life with pleasure, i.e., maximum delight,” nevertheless “the question remains as to where such delight lies.”<sup>75</sup> In other words: What *kinds* of pleasures primarily constitute a felicitous life and by what *means* do we experience them? Hobbes’s answer starkly diverges from his ancient hedonist predecessors. The Epicureans took felicity to consist in the “static” pleasures of being pain-free in virtue of having satisfied or relinquished one’s desires. They also placed greater emphasis on psychic pleasures than somatic pleasures; indeed, on the Ciceronian interpretation, static pleasures just are mental pleasures. The Cyrenaics, by contrast, located felicity in the somatic pleasures of satisfying one’s desires. Hobbes rejected both accounts.

The first point Hobbes made in response to his question is that felicity primarily consists in the anticipatory pleasures of mind that accompany desire, rather than in the pleasures of fruition that accompany sensory perception: “as nothing can be desired unless desired with regard to the future [*sub ratione future*], it is necessary to place felicity in the appetite for future good.” Why is felicity not, as the Cyrenaics taught, composed of sensory pleasures of fruition concerning the present? Hobbes answered by echoing the Epicurean view that such pleasures are too fleeting to contribute intrinsically to an agent’s *ongoing* experience: they last only so long as one’s senses are being titillated and one’s desire is *being satisfied*.

It might be asked whether the present does not please while one enjoys [*fruimur*] it. Certainly it pleases, moreover it may seem that felicity consists in present enjoyment [*fruitione*], rather than in appetite for the future; it is therefore necessary to consider that the delight which is aroused by present things is brief, in fact we are said to be enjoying them only so long as we sense [*sentimus*] them, i.e., so long as the delightful object [*objectum jucundum*] acts on our sensory organs in the very action that delights, which time in every variety of pleasure is short.<sup>76</sup>

Moreover, like the Epicureans, Hobbes added that, beyond being fleeting, pleasures of sense also tend to give way to pain: “those pleasures are usually called sensual in which the object’s action, which was at first delightful

<sup>74</sup> AW 38.8; 38.9.    <sup>75</sup> AW 38.8.    <sup>76</sup> AW 38.5; see 38.8.

[*jucunda*], on account of its lingering and of the feeling of alteration presently becomes painful [*molesta*].”<sup>77</sup> Hobbes proceeded to argue that even if a pleasure begins as a sensory pleasure of fruition, it can persist only as a pleasure of mind; any subsequent pleasures of fruition arise from perceiving the object of sense while we satisfy consecutive appetites:

if we consider the present to be not time, but an instance, we shall confess that a pleasure of sense, if it persists, is no longer in the sense, but in the memory. What is in the memory, does not please as something past, except insofar as the past events are signs of similar pleasures expected in the future. But the future does not exist, except in the imagination of we who suppose that the past is connected to the present as what follows [the past]; but what follows the present, we call the future. Hence fruition [*fruitio*] itself is nothing, in the very object we enjoy [*fruimur*], other than the latest appetite arising from the contemplation of the object’s parts [*orta à partium eius contemplatione appetitio ulterior*]; hence it remains true that the basis of good [*rationem boni*], and hence of felicity, consists in appetite.<sup>78</sup>

If felicity requires a greater amount of ongoing mental pleasures relative to pains, then Hobbes’s second point was that the amount of ongoing joy relative to grief is maximized when *desires* combine with the *hope* of satisfying them – rather than with the fear of frustrating them. For such fear would offset anticipatory pleasure by inflicting anticipatory pain. This amounts to rejecting the Epicurean dictum that unfulfilled desire is intrinsically painful: whether the presence of desire is more painful than pleasant overall depends on how confident or fearful one is about its potential satisfaction. His third point was that desire combines with hope when one has experienced the repeated satisfaction of one’s desires via a constant *accumulation* of desired goods. Felicity

consists in the progress of appetite from an acquired good to another good to be acquired . . . [but] There is no felicity in the appetite or acquisition of goods, if a greater or equal fear of losing them perpetually accompanies the appetite and hope of acquisition. For the delight [*Iucunditas*] that lies in hope is offset by the anxiety that lies in fear. It is therefore necessary for felicity to combine the hope of acquiring with the hope of conserving; and so consequently felicity consists in the progress of acquiring, such that the new acquisitions seem to protect the old.<sup>79</sup>

Thus Hobbes shared the Epicurean concern with the elimination of mental anxiety, but in rejecting the view that desire is inherently ridden with anxiety, he offered a solution sharply contrasting with the Epicurean

<sup>77</sup> AW 38.5.    <sup>78</sup> AW 38.5.    <sup>79</sup> AW 38.6.

one. In the famous opening of book 2 of *De Rerum Natura* – to which Hobbes explicitly referred in *Anti-White* – Lucretius, like his master Epicurus, had portrayed the pursuit of power as noxious to felicity.<sup>80</sup> Hobbes, by contrast, made felicity depend on it. His fourth point was that awareness of continual success in having satisfied one's desires spawns hope only because it makes one aware of one's *power* to satisfy desires:

consider that passion which we call joy, or delight of the mind, which differs from those stimulations of the body and organs that constitute sensual pleasures. Whatever is troublesome [*molestum*] for the mind, is troublesome because it lowers one's high sense of one's own power. And since power [*potentia*] concerns the future, to determine one's judgement of that power, by which all goods are acquired by us, is to determine hope.<sup>81</sup>

Felicity originates in awareness of one's power to satisfy desires in the future: "since not only acquiring, but also protecting one's gains, are a power, felicity will be the perpetual progress of appetite and hope from lesser to greater power."<sup>82</sup>

The upshot is that, although the subjective satisfaction of desires and the accompanying pleasures of fruition do not intrinsically contribute much to ongoing felicity – their contribution is too fleeting – they do play a significant *instrumental* role, by fostering anticipatory pleasures: "felicity consists not in sensory enjoyment [*fruitione sensibili*], but in delight of imagination [*jucunda imaginatione*] of progressing from the enjoyment [*fruitione*] of one good thing to the enjoyment [*fruitionem*] of another. This however is to be understood such that the progression pleases not as a progression [of fruitions], but as a proof of one's own virtue, & excellence, which are the progression's cause."<sup>83</sup> The point of the cautionary second sentence is that, whereas felicity primarily *consists* in ongoing delight of the imagination, a succession of pleasures of fruition plays a primarily instrumental role in promoting felicity, by fostering hope and hence net anticipatory pleasure. Hobbes's argument in *Anti-White* therefore makes crystal clear why when in later works he suggested that felicity relates to appetites, he was not thereby proclaiming a new, desire-fulfilment theory of the good. The desire-fulfilment theory was not on his radar. It occurs to modern readers only because of a failure to appreciate the relation Hobbes posited between appetite, pleasure, and felicity: he associated having desires with felicity insofar as they are accompanied by anticipatory pleasures; and he associated the *subjective* satisfaction of desire

<sup>80</sup> DRN II, lines 1–61; AW 38.16.

<sup>81</sup> AW 38.7.

<sup>82</sup> AW 38.6.

<sup>83</sup> AW 38.8.

with felicity instrumentally, insofar as it engenders hope of further satisfaction.

Hobbes rounded out this hedonist account by advancing the even starker thesis that *all* pleasures of mind ultimately reduce to a form of “glory,” which was Hobbes’s name for the anticipatory joy that accompanies contemplating one’s own power:

if in fact every mental grief [*dolor animi*] lies in recollecting or feigning [*fictione*] one’s own weakness, then every mental delight [*animi jucunditas*] consists in the recollection, or at least feigned imagination [*fictâ imaginatione*], of one’s own power, or excellence. Therefore joy, or delight of the mind [*delectatio animi*], is nothing but a kind of triumph of the mind, whether internal glory, or glorying over one’s own power & excellence with respect to another to whom one compares oneself.<sup>84</sup>

This was a recurring theme of Hobbes’s psychology: the assertion that all mental pleasures are reducible to glory appears in all three political treatises.<sup>85</sup> It is also firmly grounded in the basic assumption behind Hobbes’s classification of pleasures: that mental pleasures are parasitic on sensual pleasures. Mental pleasure ultimately consists in anticipating the satisfaction of desire, which anticipation depends on contemplating one’s power to satisfy it.

It is therefore no surprise that Hobbes tightly linked felicity to glory: if pleasures of mind amount to the joy of glorying, then felicity consists in the maximum ongoing experience of glory (relative to dejection of mind). Yet because felicity requires ongoing, and not just fleeting pleasures, Hobbes insisted that it requires *true* glory grounded in a true record of success:

Now if glory be such that it springs from assessing one’s power on the basis of previous deeds, they bring about hope, because he who has done, seems to have the power to do again. Therefore such a self-assessment gives rise to diligence [*industria*], & frequent success, through a true and just assessment of power; moreover one success causes another, thanks to the new power secured with each success; and this continuous manner of successes, together with a reason for hope if they persist, is called felicity.<sup>86</sup>

Hobbes contrasted this true glory with two forms of vainglory, not grounded in true successes, and neither of which ordinarily conduces to

<sup>84</sup> AW 38.8.

<sup>85</sup> EL 27.3; 8.8, 8.6; DCv 1.2; 1.5; 15.13; L 17.8: 258. L 17.8 speaks against Slomp’s (2007: 184–185) thesis that Hobbes changed his mind in *Leviathan*. Slomp’s reading is facilitated by the mistaken claim that Hobbes earlier held that “*all* human passions derive from the pleasure or displeasure” of glory. But Hobbes’s thesis concerned only *mental* pleasures. Cf. Slomp (1998).

<sup>86</sup> AW 38.8.

felicity. One kind of vainglory arises from the fictional imagination of one's power, which, because it does not engender any hope of future success, produces a merely fleeting pleasure.<sup>87</sup> The other kind arises from being falsely persuaded of one's power on the basis of flattery from others, which ordinarily produces a merely fleeting pleasure because, although it initially fosters hope, the hope is short lived: hope is immediately dashed by failures that expose one's feebleness.<sup>88</sup>

In short, felicity consists primarily in pleasures of mind, not sense; pleasures of mind are anticipatory pleasures, not fruition; pleasures of mind thrive relative to grief when desires combine with hope of satisfying them; hope depends on awareness of repeated success in having satisfied one's desires; and repeated success spawns hope and confidence or true glory because it generates awareness of one's power to satisfy desires.

To all of this it might be objected that *Anti-White*, which Hobbes never published, does not represent his considered view and hence provides an unreliable basis for interpreting Hobbes's theory of the good. We should turn instead, it might be argued, to his published works, in which Hobbes is frequently taken to have *denied* that there is any supreme good or objective criterion of felicity. It is sometimes thought that, rather than a *summum bonum*, Hobbes posited only a *summum malum*, namely death, such that survival, rather than contentment or felicity, is the proper object of desire. But the objection is groundless. Although Hobbes's published writings do not lay out his theory of felicity in the sustained way we find in *Anti-White*, they strongly reiterate its basic doctrine: a systematic perusal of Hobbes's works demonstrates that the theory's basic elements are present, even if somewhat scattered, throughout his published political texts. *Anti-White* is simply the text that most systematically stitches together the elements of the theory of the good undergirding Hobbes's political philosophy. The objection may get things backwards: perhaps we find the most systematic exposition of Hobbes's theory here precisely because he did *not* intend it for publication. Hobbes's presentation of his theory of the good in his published works does not follow the structure of justification underlying the theory; it follows Hobbes's didactic purpose. Although Hobbes conceived of felicity as the ultimate good grounding the value of all other goods, it is not the proper *aim* of all valuable action. To live a felicitous life

<sup>87</sup> "But if someone assesses his power on the basis of his own fiction (for anyone can attribute to himself, by means of a fiction of fancy [*per fictionem phantasticam*], whatever actions contrived [*fictas*] by the poets), no hope for future success arises, but a passing & vain delight [*delectatio*] only, on account of which it is called vain glory, which in no way conduces to felicity." AW 38.8.

<sup>88</sup> AW 38.8. Cf. EL 9.1; L 6.39–41: 88; II.11–12: 154–156.

one must desire and pursue other objects as well. Indeed, it was a fundamental tenet of his political philosophy that, given the human condition, felicity can be realized only by directly desiring and pursuing the conditions of *survival*. Although the criteria of a felicitous life in Hobbes's theory of *value* are maximum and ongoing net pleasure, individuals' affective and practical reasons are not necessarily, in Hobbes's theory of normative *reasons*, directly to aim for maximum, ongoing pleasure or felicity. Under conditions of uncertainty, the most reasonable course of action may involve a disaster-avoidance rather than distributed-utility-maximizing strategy.<sup>89</sup> This is why in his published works he rhetorically subordinated the language of "contentment" to the language of "self-preservation," even as he periodically reminded his readers that – as we saw in [Chapter 3](#) – self-preservation encompasses not just survival but also "contentment" or "commodious living."

Hobbes emphasized the hedonist and processual character of felicity as early as *Elements*: by felicity "we mean continual delight"; it "consisteth not in having prospered, but in prospering,"<sup>90</sup> i.e., not in fruition or a *state*, but in an ongoing process of repeated success spawning hopeful anticipation. Hobbes sharply contrasted his own anticipatory theory of felicity, as a progression from one appetite and anticipatory pleasure to another, to the ancient doctrine according to which felicity lies in the *end* or complete *satisfaction* of desire: "But for an utmost ende in which the Ancient Philosophers have placed Felicity . . . there is noe such thing in this world, nor way to it . . . for while we live we have desires, and desire presupposeth a farther Ende."<sup>91</sup> In *De Cive*, Hobbes's more political focus was restricted to individuals' "civil felicity" (*felicitatem civilem*).<sup>92</sup> But in *Leviathan* he returned to characterizing felicity in general terms. Again rejecting the view that felicity consists in the tranquility of mind that was supposed to come with the end of desire, he reiterated that

the Felicity of this life, consisteth not in the repose of a mind satisfied. For there is no such *Finis ultimus*, (utmost ayme,) nor *Summum Bonum*, (greatest Good,) as is spoken of in the Books of the old Morall Philosophers. Nor can a man any more live, whose Desires are at an end, than he, whose Senses and Imaginations are at a stand. Felicity is a continuall progresse of the desire, from one object to another; the attaining of the former, being still but the way to the later. The cause whereof is, That the object of mans desire, is not to enjoy [LL: *fruatur*] once onely, and for one instant of time; but to

<sup>89</sup> Kavka (1986: 202–207). But the normativity of disaster-avoidance would still be conditional on and derived from its ability to serve expected-distributed-utility-maximization.

<sup>90</sup> EL 9.7.   <sup>91</sup> EL 7.6.   <sup>92</sup> DCv 13.6; cf. 3.22.

assure for ever, the way of his future desire [LL: *sed ut fruitionem suam securam in futuram reddat* (but to render his enjoyment secure in the future)].<sup>93</sup>

Neither the passage from *Elements*, nor this one from *Leviathan*, nor the equivalent one in *De Homine*<sup>94</sup> says, as Hobbes has often been taken to say, that there is no *summum bonum*. It is true that, because individuals take pleasure in different objects, Hobbes denied any shared content beyond ongoing pleasure to the *summum bonum*; but he nevertheless insisted that the *summum bonum* just is ongoing pleasure. These passages say that the supreme and ultimate good for all humans consists not in the fruition or realization of some end – not in some state – but in a continual *process* of having one hopeful desire and hence anticipatory pleasure after another.<sup>95</sup> They reiterate the fundamentally hedonist, processual, and anticipatory character of felicity, which Hobbes again explicitly linked to leading “a contented life.”<sup>96</sup>

Earlier, in [chapter 6](#) of *Leviathan*, Hobbes had characterized felicity as “*Continual Success* in obtaining those things which a man from time to time desireth, that is to say, continuall prospering.”<sup>97</sup> The apparent contrast with *Elements*’ formulation, which is in terms of “continual delight” rather than success and desire, has led some to read *Leviathan*’s [chapter 6](#) passage, unlike the [chapter 11](#) passage, as reflecting a desire-fulfilment theory of well-being.<sup>98</sup> We can now see why it does no such thing. Desire satisfaction *does* play a role in Hobbes’s hedonist account, but only when one is subjectively aware of it, and even here subjective desire satisfaction plays a primarily instrumental role, by giving rise to self-confidence or true glory and hence to the hopeful anticipation of ongoing *future* satisfaction.

Hobbes confirmed *Leviathan*’s commitment to a hedonist, rather than desire-fulfilment, theory in his rare treatment, later in [chapter 11](#), of a case that distinguishes the two theories. Discussing the “Desire for Fame after death,” Hobbes wrote that “though after death, there be no sense [i.e., sensory perception] of the praise given us on Earth,” so that it can yield no pleasure of satisfaction,

yet is not such Fame vain; because men have a present delight therein, from the foresight of it, and of the benefit that may redound thereby to their

<sup>93</sup> L II.1: 150–151.    <sup>94</sup> DH II.15.

<sup>95</sup> The ancients arguably did not understand felicity as an “end” in this sense either (Annas 1993: 34–36), and it may be that Hobbes was responding more to ancient ethics in translation.

<sup>96</sup> L II.1: 150.    <sup>97</sup> L 6,58: 96.    <sup>98</sup> McNeilly (1968: 129–136).

posterity: which though they now see not, yet they imagine; and any thing that is pleasure in the sense, the same also is pleasure in the imagination.<sup>99</sup>

If the desire for post-mortem fame is not “vain” only because having such a desire is accompanied by anticipatory pleasure, that is because its contribution to felicity intrinsically lies in the pleasure it produces, rather than in whether the desire is actually ever satisfied.<sup>100</sup>

#### 4.4 Two Complications for the Meaning of ‘Good’: Prescriptively Subversive Circumstances

Hobbes held that when ‘good’ is used in its customary, pre-scientific sense, it names what agents presently desire and find pleasant. The customary sense, we might say, is Cyrenaic. By contrast, when the term is used in the reforming sense that Hobbes proposed on the basis of his hedonism, it names something else: what conduces to felicity. His scientific sense, we might say, is eudaimonist.

Given that the word ‘good’ in these two different senses does not name the same objects, we might ask what the customary and scientific meanings have in common, i.e., why Hobbes’s scientific definition is not just of a different concept altogether. The answer lies in the fact that for Hobbes the meaning of ‘good’ cannot be reduced – whether in its customary or scientific sense – to the descriptive or indicative role of attributing a property to objects. Hobbes claimed that although “generally all Passions may be expressed *Indicatively*,” as when we report on the desires we have, “The language of Desire, and Aversion, is *Imperative*” as well.<sup>101</sup> To call something good is not just to *assert* that we desire it or that it has some property, but also to “praise” or commend it, and hence to *prescribe* favouring it in some way.<sup>102</sup> The prescriptive force inherent to the meaning of ‘good’ is the thread running through both the customary and reforming definitions: the reforming definition retains the customary term’s prescriptive meaning, but changes its descriptive or indicative meaning – so that it can be used to prescribe not what agents presently desire or find pleasant, but what is conducive to their felicity.

The fact that someone calls something good and hence prescribes favouring it does not *ipso facto* give anyone a normative reason to do, desire, or otherwise favour anything. And just as prescriptions can be non-

<sup>99</sup> L II.6: 152. <sup>100</sup> DH 12.8. <sup>101</sup> L 6.55: 94.

<sup>102</sup> L 6.59: 96. Cf. DCv 3.31. The imperative or prescriptive language of desire may take three forms: command, counsel, or prayer. L 6.55: 94.

normative, so too might the term 'good'. There are therefore two ways to read Hobbes's account of the customary meaning of 'good'. The customary meaning might have indicative and prescriptive, but not normative, semantic force. On this interpretation, when a Presbyterian minister extols the virtues or goodness of rebellion against the king,<sup>103</sup> he is (a) signifying his present desire that his parishioners rebel, as well as (b) commending and hence prescribing rebellion. But he need not be making any genuinely normative claims. However, on Hobbes's account of counsel, counsellors purport to ground their prescriptions in normative reasons.<sup>104</sup> If Hobbes thought that 'good' is used, even in its customary sense, to give counsel – as he clearly did – then he recognized that it is used not just prescriptively, but normatively as well. When the minister extols the goodness of rebellion, he is also implicitly (c) claiming that upon reasoning rightly his parishioners too would desire to rebel, and this on the basis of implicitly (d) claiming they have a normative reason to rebel. This normative reason is supposed to be the basis for his calling rebellion good and prescribing it.

Yet Hobbes clearly thought that, even though 'good' in its customary sense is often used with normative semantic force, the objects customarily named good frequently fail to be genuinely normative: that something is "good" for agents in the customary sense that they presently desire it and hence find it pleasant does not *ipso facto* give them a normative reason to favour it. The normative claims made in using 'good' in the customary sense are frequently false: the customary meaning's descriptive component corresponds to agents' apparent, not true, good.

Hobbes clearly also meant for calling something good in his own reforming sense to be a way of saying there is a normative reason to favour it. For according to Hobbes's substantive prudentialist theory, goodness is a genuinely normative property. Hobbes's foundational normative principle concerns "that Reason, which dictateth to every man his own good."<sup>105</sup> No surprise if the reforming, scientific definition of 'good' proposed by Hobbes reflects not merely the prescriptive character of the term, but the inherently normative character of goodness. 'Good' in its scientific sense is supposed to have not just prescriptive but also genuine normative semantic force.

There is a significant complication concerning instrumental goods, however. On the one hand, it seems that according to Hobbes's hedonist theory of value, the property of goodness pertains to what will *actually* increase one's overall, ongoing pleasure relative to pain. Thus Hobbes

<sup>103</sup> B 2: 231; 4: 323.   <sup>104</sup> L 25.1–2: 398; 25.15: 410.   <sup>105</sup> L 15.4: 222.

characterized an agent's *bonum verum* as the total chain of actual consequences involving greater net ongoing pleasure.<sup>106</sup> And in his debate with the Foole, he appeared to concede that a violation of covenant that would actually, *ex post*, end up installing a subject as sovereign is (instrumentally) good for the subject: "some accident which he [the subject] could not expect, arriving, may turn it [his action] to his benefit." On the other hand, according to Hobbes's prudentialist theory of affective and practical reasons, one has a normative reason to favour only what one can reasonably *know* or *foresee* will (tend to) enhance one's felicity: as we have seen, Hobbes relativized normative affective and practical reasons to the epistemically accessible evidence. Hobbes denied a subject has an *ex ante* normative reason to violate his covenant – even if doing so would actually end up installing him as sovereign – on the grounds that he could not reasonably know or foresee the actual outcome. He could reasonably foresee only that such an action "tendeth to his own destruction."<sup>107</sup>

A gap therefore appears in Hobbes's ethics between his substantive theory of the *good* and his theory of *reasons*: the latter but not the former is relativized to the epistemically accessible evidence. The question is whether the normative semantic force of Hobbes's reforming definition of 'good' is supposed to reflect his epistemically non-relativized theory of the good, or instead his epistemically relativized theory of affective and practical reasons. Of course even to be able to ask this question requires using 'good' (in the phrase 'theory of the good' in the previous sentence) in the former, epistemically non-relativized sense, but this use occurs within the metalanguage from within which one considers how to fix the scientific meaning of 'good'; it does not itself count as fixing the scientific meaning of 'good'. My thesis is that, in the instrumental case, Hobbes's reforming definition is supposed to reflect his theory of affective and practical reasons, rather than his theory of value. This thesis may seem counterintuitive but it cuts to the heart of one of the most significant and original aspects of Hobbes's moral and political philosophy.

Consider the proposal to define 'good' scientifically in terms of means that turn out *actually* to increase felicity. That this was not Hobbes's proposal becomes clear in light of the fact that the prescriptions of natural law are the *paradigmatic* instance of what his science of ethics calls "good"<sup>108</sup> – and recall that the relevant means to self-preservation prescribed by natural law are not the means that actually promote self-preservation, but those that one can reasonably know or foresee do so.

<sup>106</sup> AW 30.25. Cf. DH II.5.    <sup>107</sup> L 15.5: 222.    <sup>108</sup> L 15.40: 242.

Take the non-social means of self-preservation. Hobbes mentioned these non-social means in order to set them aside as not “pertinent enough” to his political subject matter. His list of natural laws focusses on those prescribing the *social* means of self-preservation – the “means of the conservation of men in multitudes; and which onely concern the doctrine of Civill Society.” But he conceded that the set of precepts forbidding those “things tending to the destruction of particular men,” may “also be reckoned amongst . . . the Law of Nature.”<sup>109</sup> One such law, noted in both *De Cive* and *Leviathan*, forbids drunkenness, on the grounds that “intemperance tends [*tendit*] to sicknesse and death.”<sup>110</sup> Hobbes was not here stupidly claiming that intoxication in every instance delivers sickness and death. (Aubrey reported that Hobbes was “temperate, both as to wine and women,” but “had been in excesse in his life, a hundred times.”<sup>111</sup>) Hobbes was claiming instead that intoxication *tends* to painful and self-destructive outcomes. Sometimes, partaking of the pleasures of drink may actually enhance one’s net pleasure. But drunkenness is nevertheless against natural law, according to Hobbes, and we can see why by analogy with his reply to the Foole: given the self-destructive “tendency” of intemperance, one cannot reasonably know or foresee that this instance of drunkenness will in actual fact turn out to be more pleasurable in its overall consequences. Temperance is therefore properly called a *virtue* and an (instrumental) *good* according to Hobbes’s science of ethics. So it would seem that, for the means to felicity, ‘good’ is defined scientifically relative to the epistemically accessible evidence, not on the basis of what will actually have the best consequences. Its scientific definition is to be fixed according to Hobbes’s theory reasons, not his substantive theory of the good.

There is a further potential complication, however. What if calling temperance “good” and drunkenness “evil” would, via some odd psychological quirk, increase my disposition to abuse alcohol and therefore tend to damage my health? And what if I can reasonably foresee this? (Perhaps I resent the moralizing terminology of good and evil and act from spite, or perhaps I get a thrill from transgressing taboos. Whatever the motivation, in the heat of the moment I tend to do what has been labelled “evil.”) The possibility that the particular vocabulary through which ethical notions are expressed may affect one’s ability to live up to their standards had already been broached by Cicero in *De Finibus*. He there characterized

<sup>109</sup> L 15.34: 238.

<sup>110</sup> DCv 3.32; L 15.34: 238. In *De Cive*, but not *Leviathan*, Hobbes also listed temperance amongst the social laws of nature. DCv 3.25.

<sup>111</sup> BL I: 350.

Zeno and his Stoic followers as claiming that virtue is the only intrinsic good constitutive of felicity; they therefore denied the Aristotelian view that there are non-moral, natural goods, such as bodily health, also constitutive of felicity. But since the Stoics maintained that bodily health, while not a *bonum*, is nevertheless *aestimabile*, “worthy of selection” (*dignum aestimatione*), or to be “preferred” (*praepositum*),<sup>112</sup> Cicero accused them of dressing up the old Aristotelian ethics in new terminology to express the same substantive doctrine. “Since there is agreement in substance,” Cicero asked, “why not prefer the standard terminology?” The only justification for the new Stoic terminology, Cicero suggested, would be if by calling natural goods “preferred rather than good,” and by refraining from calling natural evils “evil,” one would thereby foster greater virtue. Cicero went on to suggest that the Stoic terminology does no such thing; it may even make non-moral goods seem more attractive.<sup>113</sup> But his insight is that the vocabulary in which ethical notions are expressed may facilitate or hinder their ability effectively to guide a human life – which, for Cicero, is the very point of ethical theory.<sup>114</sup>

Thus the complication is that in some circumstances the choice of what to call “good” and what “evil” might foreseeably affect behaviour in subversive ways. If I call “good” that which tends to maximize my felicity, doing so might in effect tend to *minimize* my felicity. The point of prescribing something to people, including to oneself, is to bring something about; prescribing is not the only way to further some end, but it is one way – a way that relies on linguistic uptake. In *prescriptively subversive circumstances*, however, prescribing something by calling it good subverts the very point of the prescription. I may have (a) an epistemic reason to believe that temperance is instrumentally good for me, an affective reason to endeavour or be disposed to it, and a practical reason to practise it, but under prescriptively subversive circumstances I also have (b) a (prudential) practical reason *not* to call temperance (instrumentally) “good.”

Hobbes gave no indication whether prescriptively subversive circumstances ever obtain for the non-social means to self-preservation. But one of the central claims of his moral and political philosophy is that they frequently do arise for the *social* means to self-preservation.

Consider rules for regulating my interaction with you that would, if we *both* abided by them, promote our self-preservation – our ongoing survival and felicity – by promoting peace between us. Take Hobbes’s sixteenth law of nature, which prescribes that parties faced with a controversy submit “to

<sup>112</sup> DF III.20–59.    <sup>113</sup> DF IV.72–73. Cf. NE 10.1.    <sup>114</sup> DF 4.52.

the judgement of an Arbitrator,” along with the eighteenth law, which prescribes impartial arbitrators.<sup>115</sup> According to Hobbes, we can each reasonably know that if we both follow these rules in our interactions, doing so will tend to enhance the felicity of each. Now imagine that, based on the evidence available to us both, we could reasonably foresee that, amongst all potential arbitrators, Taraneh would be the most unbiased. Under these circumstances we can both reasonably know or foresee that setting up an unbiased arbitrator to settle our controversies, installing Taraneh as arbitrator, and abiding by her judgements would tend to enhance our felicity. You have reasoned rightly, and therefore know these natural laws; given your extensive experience and epistemic prudence, you apply these laws so as to pick out Taraneh as the best arbitrator; and given your affective/practical prudence, you are disposed to abide by her judgements if I do so as well.

It may nevertheless turn out that you calling the sixteenth and eighteenth laws of nature or Taraneh’s installation as arbitrator “good” would diminish our felicity. Such circumstances might arise if I *disagree* with you about which rule or which of its applications would be most beneficial – for example, if I have reasoned wrongly, and so fail to know the sixteenth and eighteenth laws, or if, given my inexperience, I fail to recognize Taraneh’s impartiality and favour instead the biased Aram. Under these circumstances, I might call Taraneh’s installation “evil.” If you insist on calling it “good,” then we face a new controversy, not a way to resolve controversy. It remains true, *ex hypothesi*, that installing Taraneh as an arbitrator to which we both defer would be good for both of us. But since we disagree about the matter, your *calling* Taraneh’s effective installation “good” would tend to *damage* our felicity if our expressed disagreement instigated conflict and war between us. And if it could be reasonably known or foreseen that such disagreement instigates war, then a gap would arise between (a) what actions we both have an epistemic reason to believe are good, an affective reason to desire, and a practical reason to undertake, on the one hand, and (b) what actions you have a (prudential) practical reason to call “good,” on the other. Under these circumstances, there would be actions that agents have reasons to undertake but that they have reasons to refrain from calling “good.” The practical point of calling Taraneh’s installation good and hence prescribing it is to promote peace and felicity; but in calling it good you would foreseeably entrench the state of war.

<sup>115</sup> L 15,30–32: 238, italics removed.

It is one of the central claims of Hobbes's moral and political philosophy that such *prescriptively subversive circumstances predominate in the state of nature*. The reason they predominate is that (1) the state of nature is plagued by widespread, expressed normative disagreements and that (2) expressed disagreement – especially normative disagreement – is a fundamental cause of war. What is more, we can reasonably know this: that expressed disagreement instigates war is what Hobbes's science is supposed to demonstrate to any reasoning agent.<sup>116</sup> Hobbes had already advanced this disagreement theory of war with crystal clarity in *Elements*: even though the laws of nature prescribe the relevant means to peace, and even though these prescriptions are discoverable by reasoning, Hobbes asserted that disputes nevertheless break out in the state of nature over “what is to be called Right, what Good, what, vertue,” and that such disputes instigate war: “In the State of nature, where everyman is his owne Judge, and differeth from other, concerning the names and appellations of things; and from those differences, arise quarrells, and breach of Peace, it was necessary, there should be a Common measure of all things, that might fall in Controversie.” No such “Common measure” can be effectively established within the confines of the state of nature.<sup>117</sup>

Hobbes elaborated on this in *De Cive* by distinguishing between disagreement over two kinds of object. Because “the appetites of men are diverse, according to their diverse tempers, customes, opinions,” we firstly disagree over the evaluative names of “those things we perceive by sense, as by tasting, touching, smelling,” that is, over the present objects in which we take a sensory pleasure of fruition. We cannot overcome disagreement in the state of nature over these present, sensory goods. Thus in order to reach agreement over the use of evaluative words, we must focus instead on the *future* goods at stake in our social interaction.<sup>118</sup>

It is true that in the state of nature we also disagree, secondly, over the evaluative names of those things “which pertain to the common actions of life” and their future consequences, i.e., over the future results of social interaction in which we might take a mental pleasure of anticipation. Hobbes was especially exercised by this latter kind of disagreement, because insofar as it concerns the rules of social interaction, it is the more explosive source of conflict. But Hobbes thought that, at a very general level, disagreement over what types of social dispositions and actions to call “good” or “virtuous” could be overcome within the state of nature itself. Precisely because the state of war visits death and misery upon its

<sup>116</sup> Abizadeh (2011). <sup>117</sup> EL 29.8. <sup>118</sup> DCv 3.31.

inhabitants, “All men easily acknowledge this state, as long as they are in it, to be evill, and by consequence that *Peace* is good. They therefore who could not agree concerning a present, doe agree concerning a future Good.” Having agreed that peace is good, we might even be able to agree on which *nominal types* of social disposition and action are instrumentally good for securing peace, because “Reason declaring Peace to be good, it followes by the same reason, that all the necessary means to Peace be good also, and therefore, that *Modesty, Equity, Trust, Humanity, Mercy . . .* are *good Manners*, or habits, (that is) *Vertues*.”<sup>119</sup> Reasoning is able to produce agreement not only on the fundamental law of nature prescribing peace, but also on secondary, derivative laws of nature.

We nevertheless face a second level of disagreement, which is insurmountable so long as we remain in the state of nature. Even if we agree on the goodness of the fundamental law, and nominally agree about the virtues prescribed by the secondary laws derived from it, we will continue to disagree about their content and hence application:

because men cannot put off this same irrationall appetite, whereby they greedily prefer the present good . . . before the future, it happens, that though all men doe consent in the commendation of the foresaid vertues, yet they dissent still concerning their Nature, to wit, in what each of them doth consist. For as oft as anothers good action displeaseth any man, that action hath the name given of some neighbouring vice; likewise wickedness, which pleases them, is ever entituled to some Vertue. Whence it comes to passe that the same Action is prais'd by these, and call'd Vertue, and dispraised [*culpari*] by those, and termed vice.<sup>120</sup>

Hobbes reiterated this analysis in *Leviathan*. He again argued that people disagree at two levels about what to call good. They differ, firstly, over the present objects of sensory pleasure: “divers men, differ . . . in their Judgement on the senses, of what is pleasant, and unpleasant to the tast, smell, hearing, touch, and sight.” It is true that, once we turn to future results in which we can take anticipatory pleasure, we are able to discover via reasoning a common good consisting in peace and the types of social interaction that promote it: because “the condition of meer Nature . . . is a condition of War, . . . consequently all men agree on this, that Peace is Good, and therefore also the way or means of Peace.” But notwithstanding the possibility of agreement at this general level, “so long . . . as private Appetite is the measure of Good, and Evill” – as it is in the state of nature – “divers men” will, secondly, differ

<sup>119</sup> DCv 3.31.    <sup>120</sup> DCv 3.32; cf. 14.17.

in their Judgement . . . also of what is conformable, or disagreeable to Reason, in the actions of common life. Nay, the same man, in divers times, differs from himselfe; and one time praiseth, that is, calleth Good, what another time he dispraiseth, and calleth Evil: From whence arise Disputes, Controversies, and at last War.<sup>121</sup>

The *actual reasoning* of individuals will not eliminate disagreement over which specific dispositions and actions to call good.

#### 4.5 Prescriptively Self-Fulfilling Circumstances

How did Hobbes propose to settle the scientific definition of ‘good’ in light of such apparently intractable disagreement and hence prescriptively subversive circumstances? Assume there exist particular dispositions or actions that would best promote peace if enough people in the state of nature had those dispositions or undertook those actions. Assume further that this is a language-independent, epistemically accessible fact of the matter. It might initially seem that such dispositions and actions would be prime candidates for what Hobbes was proposing to call “good” in his scientific sense. But the problem, under prescriptively subversive circumstances, is this: if normative disagreement in the state of nature about particular dispositions and actions is insurmountable, and given that controversy triggers war, then even if such candidates existed, if one insisted on calling any such candidates “good,” one’s doing so would end up contributing to the perpetuation of war. Since this eventuality could be reasonably known or foreseen, a gap would obtain between one’s epistemic reasons for believing that such candidates – assuming they exist – are good, on the one hand, and one’s practical reasons for calling them “good,” on the other. If *all* such candidates would be subject to disagreement, then since each of us has a prudential practical reason to restrict normative language to those things over which we could reasonably hope to agree, each would have a practical reason to *refrain* from calling anything “good” – even if we had an epistemic reason to believe that some social means to peace are available. Hence Hobbes’s scientific theory appears hospitable to two, potentially conflicting standards for how to reform and fix the definition of ‘good’: according to what would, if agreed by a sufficient number, conduce to peace and hence felicity, or according to the use of language that would itself conduce to peace.

<sup>121</sup> L 15.40: 242.

Hobbes's solution to this problem – which seeks to close the gap between these two standards – ironically relies on the fact that for the means to peace the problem of disagreement cuts even deeper still. It cuts even deeper because, for Hobbes, the means to peace are not just *consequentially* but also *constitutively* social. That their consequences are social is clear: means to peace help to preserve a multitude, not just an individual. But they are also constitutively social in the sense that they consist either in (a) the separate but common dispositions or actions of a multitude or in (b) the joint actions of a multitude. The third and fourth laws of nature, for example, which prescribe abiding by covenants and expressing gratitude for favours, are constitutively social in the former way: actually adhering to covenants will promote peace only if a sufficient number of individuals do so. The sixteenth law of nature is constitutively social in the latter way: no one can unilaterally submit contending parties to voluntary arbitration.<sup>122</sup> No single individual's dispositions or actions can unilaterally establish peace amongst a multitude.

Because the means to peace are constitutively social, if normative disagreement about particular dispositions and actions were insurmountable, then even if there existed candidates that under counterfactual circumstances *would* be the means to peace, it would nevertheless turn out that none *are* the means to peace for anyone: there would not be any means to peace available to individuals in the state of nature. From the fact that some action of yours hypothetically *would*, under circumstances of sufficient agreement and coordination, promote the best outcomes for each, it does not follow that it *will* do so; if there is significant noncompliance, it might promote very bad outcomes instead. This was one of Hobbes's essential points when he argued that natural laws do not articulate universally binding practical reasons to act (*in foro externo*) in the state of nature. If renouncing your right to all things *would* promote peace only on condition that others also renounce, then your renouncing *will* actually promote peace only if a sufficient number of other people actually do or will renounce too; if they do not, then your renouncing *is not* a means to peace – it may instead be a means of self-destruction.<sup>123</sup>

It follows that an action of yours is a relevant means to peace for you only if you can reasonably know or foresee that it *will* promote peace. If you can reasonably know or foresee that your  $\phi$ -ing will promote peace only on condition that others also  $\phi$ , then your  $\phi$ -ing is a relevant means to peace for you only if you can reasonably know or foresee that the

<sup>122</sup> L 15.30: 238.    <sup>123</sup> L 14.5: 200.

hypothetical antecedent condition – a sufficient level of reciprocal compliance – will be met. If the hypothetical condition could be *made* to obtain, then such candidates could *be* means for promoting peace. But if the hypothetical condition does not presently obtain, and cannot be made to obtain in the future, and if you could reasonably know or foresee this, then you have no relevant means for promoting peace. If disagreement concerning the specific dispositions or actions on which it would be (instrumentally) good to coordinate is insurmountable, then we will not foreseeably coordinate, and so no particular dispositions or actions will count as relevant means to peace for us. The relevant, constitutively social means to peace exist for individuals only if realizing the hypothetical condition is foreseeably feasible.

This is why the social means to self-preservation in the state of nature face not merely prescriptively subversive circumstances – as non-social means might – but *deeply* subversive circumstances. Recall: calling temperance good may be prescriptively subversive for individuals seduced by the charms of transgression. But calling temperance good is not deeply subversive: if these charmed individuals call temperance good, they undermine their own preservation by disposing themselves to abuse alcohol, but temperance nevertheless remains a relevant means for self-preservation. Such individuals do not, in virtue of calling temperance good, falsify the truth that it is a relevant means to self-preservation; what undermines their self-preservation is not temperance, but calling it good. Similarly, if some candidate is the social means to peace, calling it good may trigger disagreement and therefore war, without thereby falsifying the truth that the candidate is a social means to peace. However, because the means to peace are constitutively social, for any particular individual their status as the relevant means depends on whether others will actually coordinate on those putative means. And whether the hypothetical condition will be met – i.e., whether a sufficient number of other people will act in ways that make one's present actions a relevant means for peace – may depend on whether those actions are called good. If calling some potential means good will trigger controversy that prevents coordination, then we face *deeply* subversive circumstances: the candidate turns out not to be the relevant means to peace precisely because it is praised for being such a means.

That Hobbes was alive to prescriptively subversive circumstances stems from his recognition that language is the tool of not just *observers* of but also *participants* in the world. Language can be thought of as performing two fundamentally different roles. On the one hand, it can be thought of as standing above or outside of reality. From this *external*, third-personal

perspective, it is the medium of an observer who, standing outside the chain of causes, uses language purely constatively to describe various states or events. On the other hand, language can be thought of as located within reality. From this *internal*, first-personal perspective, it is the medium of a participant who uses it not only to describe, but also to prescribe (bringing about) various states or events. From this perspective, speech stands squarely within the chain of causes. Prescriptively subversive circumstances – whether shallow or deep – arise because language, in its internal role, is itself a part of the reality that language, in its external role, is used by observers to describe: any attempt to prescribe something, on the basis that it is the causal means to realizing a valuable end, may itself affect and so undermine that basis.

Because of this internal role of language, and because the means to peace are constitutively social, what the relevant means *are* may turn out to depend, causally, on which potential means are *called* good. There may therefore be no language-independent fact of the matter as to whether specific dispositions or actions are the relevant means to peace for individuals. Consider three germane scenarios. First, if one can reasonably know or foresee that under no circumstances will there be sufficient compliance by others – regardless of what the potential means are called – then one has neither an epistemic reason to believe they are the relevant means, nor a practical reason to call them good. These are not prescriptively subversive circumstances, and there are no grounds, either epistemic or practical, for calling such candidates good: the infeasibility of the hypothetical condition defeats any proposal for calling them instrumentally good. Second, if one can reasonably know or foresee that there exist feasible conditions for sufficient compliance, but that calling the candidates “good” would tend to *prevent* the realization of the hypothetical condition by undermining sufficient compliance, then one has an epistemic reason to believe they are the relevant means, but a practical reason not to call them good. These are prescriptively subversive – indeed, deeply subversive – circumstances. Third, if presently calling some dispositions or actions “good” contributes to sufficient future compliance, and if this would *make* those candidates the best means for promoting peace, then they presently *are* the relevant means to peace; and if one could reasonably know or foresee this, then one would have both an epistemic reason to believe they are the relevant means, and a practical reason to call them good. These are not subversive but, rather, *prescriptively self-fulfilling circumstances*: in virtue of calling something good – in virtue of prescribing it as the relevant means to a valuable end – we make it (instrumentally) good.

Hobbes's fundamental insight is that the point of calling candidates for the social means to peace "good" is to contribute, by so calling them, to *making* the hypothetical condition obtain. This "solution" to the gap between the two potentially divergent standards for fixing the scientific meaning of 'good' depends on the possibility of agreeing and coordinating on some specific dispositions or actions even in the state of nature. Agreement on peace and its potential means at a general level is insufficient: if we disagree about *specific* dispositions and actions, then we will not reciprocally coordinate on them, and those dispositions and actions will not be social means to peace. Therefore, the question is whether it is possible to agree and coordinate on any specific dispositions and actions and, if so, which ones.

Hobbes allowed for limited, local, and temporary agreements in the state of nature. But his core thesis is that the only way (a) to establish a stable and general agreement about the specific dispositions and actions to call good, and (b) to ensure coordination on those actions, is to refrain from individually deciding what to call good and defer instead to a common sovereign empowered to enforce coordination on what he deems the relevant social means to peace. It "was necessary," Hobbes argued in *Elements*, "there should be a Common measure of all things, that might fall in Controversie." Although "some say" this common measure "is Right Reason," in reality right reasoning consists in an ideal, and not in anyone's actual reasonings. Because there is no "such thing to be found or knowne in rerum natura," that is,

seeing right reason, is not existent; the [actual] reason of some man, or men, must supply the place thereof. And that man, or men, is he, or they, that have the soveraign power . . . And consequently, the civill Lawes, are to all subjects the measures of their Actions, whereby to determine whether they be right or wronge, profittable, or unprofitable, virtuous, or vitious; and by them, the use, and definition of all names not agreed upon, and tending to controversie, shall be established.<sup>124</sup>

The foundation of Hobbes's political philosophy is the claim that *there is only one specific act conducive to peace on which a sufficient number of individuals in the state of nature could come to agree and coordinate*, namely, the act of establishing a common sovereign. And he assumed that in moments of crisis, those who find themselves without a sovereign will come to share his core insight. They will not only agree that peace – the

<sup>124</sup> EL 29.8.

relevant means to self-preservation – is good; they will also agree that a single sovereign power is the necessary and sufficient means to peace.

If the only state-of-nature act that can establish peace is covenanting to establish a sovereign, and if the hypothetical condition of sufficient agreement and coordination cannot be met for any other disposition or act, then does it follow that the derivative laws of nature prescribing other dispositions and actions always fail to be normative in that state? It does not. For even if the dispositions and actions prescribed by natural law are not, under circumstances of widespread non-compliance, social means to *peace* for any individual, they may nevertheless be relevant means to *individual* self-preservation. Consider the second and third laws of nature, which respectively prescribe that one (be disposed to) make covenants “*for Peace, and defence*” and to “*performe*” one’s covenants.<sup>125</sup> Take a state-of-nature covenant that is not a sovereignty covenant, but which nevertheless offers parties the prospect of mutual gain. Hobbes argued that the Foole who violates such covenants will tend to be excluded from confederacies – temporary state-of-nature alliances for mutual defence – whereas those who make and fulfil covenants will incline others to consider them as potential confederates.<sup>126</sup> Confederacies are indispensable for self-preservation in the state of nature: as long as that state persists, “there is no man can hope by his own strength, or wit, to defend himselfe from destruction, without the help of Confederates.” It is true that alliances of mutual defence are temporary and insecure means of self-preservation: they do not end the state of war.<sup>127</sup> But even if a specific action is not the social means to peace, one may nevertheless have a decisive prudential reason to undertake it if it is the relevant means to self-preservation. Moreover, Hobbes argued, if one violates the laws of nature, others will tend to exclude one not only from confederacies, but also from commonwealths: others will tend to refuse such a person “into any Society.”<sup>128</sup> Thus although my unilateral expressions of gratitude cannot be the means to *peace* in the state of nature, they are nevertheless relevant means for making sovereignty covenants *with me* in particular possible: by expressing gratitude I dispose those who presently defy natural law to become receptive to sovereignty covenants with me when the occasion presents itself in the future.

The act of establishing a sovereign therefore has a double status. Each can also reasonably know or foresee (a) that the act of establishing sovereignty will be the first *object* of widespread agreement and coordination in

<sup>125</sup> L 14.5: 200; 15.1: 220.

<sup>126</sup> L 15.5: 224.

<sup>127</sup> L 17.4–5: 256–258.

<sup>128</sup> L 15.5: 224.

the state of nature and (b) that establishing a sovereign will *produce* agreement and coordination on the other social means of peace. Sovereignty covenants therefore constitute the relevant social means to peace for each person, and make it the case that other specific dispositions and actions will subsequently *become* social means to peace. For if some candidate *would* be a social means to peace for me if and only if others coordinate on it, and if the sovereign is able to secure agreement and coordination, then in virtue of calling those dispositions and actions “good” and enforcing coordination, he *makes* them *be* the social means to peace for each subject: they *are* good because the sovereign *says* they are.<sup>129</sup> The sovereign’s say-so makes things good because in calling something good and hence prescribing it he does not merely *counsel* favouring it: he *commands* it as law and *enforces* his laws via punishments: “the whole power of prescribing the Rules” or “Civill Lawes” that determine what is “*Good, Evill, Lawfull, and Unlawfull* in the actions of Subjects” is “annexed to the Sovereigntie.”<sup>130</sup> That the sovereign’s say-so takes the form of enforced law is why it actually produces coordination – which is the hypothetical condition that must obtain in order for specific dispositions or actions to be social means to peace for anyone.

This explains why Hobbes could advance his own independent claims about which particular things are common goods (namely, the social means to peace prescribed by natural law), while at the same time claim that whether something is good or evil “in the actions of common life”<sup>131</sup> depends on the sovereign’s say-so. It is independently true that something *would be* commonly good if it were the means to peace, but whether a specific action *is* a means to peace depends on his say-so. This is why it seems the sovereign’s interpretation of the social laws of nature cannot be mistaken in a commonwealth: such laws are whatever the civil laws say they are; they are, as Hobbes put it, “a part of the Civill Law in all Commonwealths.”<sup>132</sup> Prescriptively subversive circumstances predominate in the state of nature, but within the commonwealth the sovereign establishes prescriptively self-fulfilling circumstances. With respect to the *social* means of self-preservation, Hobbes was in the final analysis proposing to fix the scientific definition of ‘good’ such that, in the commonwealth, it names whatever the sovereign says it names. This appears to close the gap between what one has an epistemic reason to believe is a relevant means to peace, and what one has a practical reason to call a common good.

<sup>129</sup> Cf. Wolin (2006). <sup>130</sup> L 18.10: 274. <sup>131</sup> L 15.40: 242. <sup>132</sup> L 26.8: 418.

Precisely because Hobbes took himself to have closed this gap for the goods he deemed most important to his political philosophy, it is not ultimately clear, in cases where the gap remains, which of the two criteria should determine the scientific meaning of 'good'. The gap between what one has an epistemic reason to believe is a relevant means to self-preservation and what one has a practical reason to call good remains for any potential means that do not enjoy prescriptively self-fulfilling circumstances. Consider again the non-social means of self-preservation. What if the sovereign prescribes regular bouts of abusive intoxication for his subjects? There are two, cross-cutting considerations here. On the one hand, since the fact that temperance is a means to self-preservation is not a constitutively social fact, the sovereign's say-so cannot *unmake* the fact that it is such a means. On the other hand, calling temperance good contrary to the sovereign's will would, by challenging sovereign power, undermine *other* (social) means to self-preservation. Once again, one would have an epistemic reason to believe that temperance is good, but a prudential reason to refrain from calling it good. The gap also remains for sovereign commands concerning things that *cannot* and so would not serve as means to peace, even should everyone coordinate on them. (Consider, at the limit, a sovereign command for all to engage in civil war.)<sup>133</sup> "For the Law-makers are men," Hobbes noted in *Questions*, "and may erre, and think that Law which they make is for the good of the people sometimes when it is not."<sup>134</sup> Such gaps are politically relevant for Hobbes, and their relevance is intensified by a further feature of his moral and political philosophy. The sovereign's say-so does not merely create reasons of the good to call something good or evil. It also creates reasons of the right for which one is accountable to others. This recalcitrant gap in Hobbes's theory of the good reflects not only the ultimate undecidability of the scientific meaning of 'good', but also the deepest structure of Hobbes's ethics: its bifurcation into two distinct and mutually irreducible dimensions of normativity.

<sup>133</sup> Hence the possibility of "iniquitous" sovereign commands. L 18.6: 270; 21.7: 330; 24.7: 390.

<sup>134</sup> Q: 146/193.



PART III

*Reasons of the Right*

Hobbes's ethics is underwritten by a fundamental distinction between two types of normative reasons, corresponding to two distinct notions of normative responsibility and blame. The first dimension of normativity concerns reasons for which we are normatively responsible in the sense that the mental states and actions to which they apply are *attributable* to us and that, should we fail to heed them, we are the proper object of censure or criticism. We are the proper object of this *critical* type of normative blame if and only if, and in virtue of the fact that, we are capable of reflectively grasping, responding to, and being guided by reasons – or, as I shall say, capable of answering to reasons. That we are capable of answering to reasons, however, does not imply we are obligated to others to respond or even listen to their criticism or advice. By contrast, the second dimension involves reasons for which we are normatively responsible not only in the sense of being open to criticism, but also of being *accountable* to others for them: should we fail to heed reasons of this kind, those to whom the reasons are owed may justifiably censure us in the stronger, juridical, and second-personal or *reactive* sense. Censure in this sense – what Hobbes called “accusation” and “condemnation” – intrinsically demands a normative response accounting for our actions and the attitudes they express – a response such as a justification, excuse, or acknowledgement of wrong and hence apology, compensation, or redress. The first, attributability dimension of normativity is treated by “ETHIQUES” or the science of “*Good, and Evil*” and comprises *reasons of the good*; the second, accountability dimension is treated by “The Science of JUST and UNJUST” and comprises *reasons of the right*.<sup>1</sup> The former corresponds to the *eudaimonistic* notion of obligation Hobbes retained from classical natural-law theory; the latter to the *juridical* notion of obligation proper, which emerged in the seventeenth century with Suárez, Grotius, and Hobbes.

<sup>1</sup> L 9: 130–131; 15.40: 242.



*Accountability and Obligations*

Hobbes's account of responsibility arose through a complicated reaction to the Christianized Aristotelian view of his scholastic contemporaries. Aristotle had taught that actions are ethically blameable only if (a) intentional or *voluntary* (*hekousios*), i.e., up to the agent and carried out knowingly, and (b) *reason-responsive*, i.e., by agents capable of decisions (*prohairesis*) resulting from rational desire (*boulēsis*) and rational deliberation (*bouleusis*). By contrast, actions compelled by force, or done through non-culpable ignorance, are involuntary and hence excusable; and beasts incapable of rational decision and deliberation are not ethically responsible or blameworthy.<sup>2</sup> For Aristotle, an action is up to oneself just in the sense that one's own desires or beliefs are its principle or beginning (*arkhē*). But when the scholastics appropriated his account, they took being under one's power to mean something more: *boulēsis* had been transformed into *voluntas* (will), and they premised ethical responsibility and blame on *free will*, i.e., the multiway power to do otherwise.<sup>3</sup> As Aquinas put it, "to praise or blame is nothing else than to impute to someone the malice or goodness of his action," where "an action is imputed to an agent, when it is in his power, so that he has *dominium* over his act." And "having dominion" requires not just being reason-responsive, but also having free will: rational beings "have *dominium* over their acts through their free will, which is the faculty of will and reason," and "Man is *dominus* of his acts, and of willing and not willing, because of his deliberate reason, which can be bent to one side or another."<sup>4</sup> Whatever their disagreements with Aquinas, Hobbes's late-scholastic contemporaries – such as Suárez and Bramhall – agreed on this basic point.

As we saw in [Chapter 2](#), Hobbes repudiated Aristotle's account of deliberation and reduced the will to the last desire before action. But in one respect he simply restored the older, pre-Christian view: he premised

<sup>2</sup> NE 3.1–5; 5.8; 7.6.    <sup>3</sup> See Pink (2011).    <sup>4</sup> ST I-II.21.2; 1.2; 109.2.

normative responsibility and blame not on metaphysical free will, but on being answerable to reasons.

### 5.1 Three Types of Responsibility and Blame: Non-Normative, Critical, and Reactive

An entity need not be an intentional agent to be *causally* responsible for something.<sup>5</sup> One may consequently be justified in “blaming” even inanimate objects in the *non-normative* sense of taking them to have caused an unwelcome outcome. One may take the rock that hit one’s head to be causally responsible for one’s headache, and therefore blame it in this non-normative sense: one blames the rock because it has caused something contrary to one’s hopes or expectations, even though the standards or precepts on which one’s expectations are premised – “Don’t hit my head!” – could not be normative for the rock. This is precisely what Hobbes took the English word ‘blame’ to mean in its most generic sense: outward “blaming” is nothing but “dispraise” or “saying the thing blamed is ill or imperfect” according to some standard or expectation whose non-fulfilment one finds displeasing. This is why “we do as much blame” inanimate things “as we do men; for we say fire hath done hurt, and the poyson hath killed a man, as well as we say the man hath done unjustly.”<sup>6</sup> Hobbes recognized that such blame may even take the form of anger, which is why in *Elements* he rejected the classical, Aristotelian definition of anger as “greife proceeding from an opinion of Contempt.” This definition “is confuted by the often experience we have, of being moved to Anger, by things inanimate, and without Sense, & consequently incapable of Contemning us.”<sup>7</sup>

Yet inanimate objects cannot be the appropriate target of *censure* and *sanctions* – such as threats or acts of revenge – aimed at providing them with incentives to comply with one’s expectations. One may *take* the rock to be causally responsible for one’s headache, and hence direct one’s anger *at* it, but one cannot *hold* the rock responsible: there is no sense in censuring it or addressing one’s anger *to* the rock. Only intentional or – as Hobbes put it – *voluntary agents*<sup>8</sup> are the appropriate target of censure and sanctions, for only they can respond to incentives. Voluntary agents are animate bodies capable of “*voluntary actions*” mediated by “thought” or “Imagination” and consequent to “deliberation” and “*Willing*.”

<sup>5</sup> Hart (2008: 214–215).

<sup>6</sup> Q: 39–40/53–54. Smart (1961) calls this “grading” blame.

<sup>7</sup> EL 9.5. See DH 12.4.

<sup>8</sup> LN §28.

The category comprises humans, but also beasts.<sup>9</sup> Thus rewards and punishments are not properly defined, as Bramhall claimed, as intrinsically deserved or “due to honest and dishonest [past] actions”; they are merely incentives directed prospectively to frame the will, as much of “brute beasts” as of “men.”<sup>10</sup>

Beasts can be *held* (causally) responsible via censure and sanctions because, despite being non-rational, they are voluntary agents who respond to incentives. Through experience they prudentially learn that certain actions provoke retaliation. Some animals even express such non-normative blame amongst themselves, to incentivize others to fulfil their expectations: “Creatures irrationall . . . have noe conception of right and wronge,” but they do have a conception of their own “pleasure and paine,” and consequently engage in a kind of “censure of one another” premised on their conception of who is causally responsible for their pain or disrupted “ease.”<sup>11</sup>

Yet although one may take beasts to be causally responsible for outcomes, blame them in a non-normative sense, and even hold them responsible by censuring and imposing sanctions expressing such blame, they cannot be *normatively* responsible for their attitudes or actions. For non-rational creatures cannot understand general precepts: the reason why “Over naturall fooles, children, or mad-men there is no Law, no more than over brute beasts” is that they lack the intellectual “means to take notice of it.”<sup>12</sup> Because they cannot understand general precepts, and hence are incapable of answering to reasons, no precepts can be normative for them: they may act imprudently, but their imprudent actions are not the proper object of normative censure. The censure addressed to or amongst animals is not premised on any normative notions such as *guilt*, *sin*, *fault*, *duty*, *obligation*, or violation of *law*. As Hobbes put it in *De Cive*, non-rational creatures such as children “are free from guilt [*culpâ vacanti*],” “because wanting the use of reason, they are exempted from all duty [*officiorum*].”<sup>13</sup> Revenge against animals expresses a *threat* – that future violations will be met with harm – but it does not express either counsel and criticism or a demand to acknowledge a norm.<sup>14</sup> If one retaliates against non-rational creatures, it is to deter them from repeating their past offence, not to prod them into acknowledging reasons or into condemning their actions on the basis of a precept or demand.

<sup>9</sup> L 6.1: 78; 6.51–54: 92–94. <sup>10</sup> Q: 132/173, 148–149/195–196. LN §14.

<sup>11</sup> EL 19.5. DCv 5.5; L 17.11: 260. <sup>12</sup> L 26.12: 422. <sup>13</sup> DCv Preface.13. See Q: 277/357.

<sup>14</sup> On human-animal relations, see EL 22.9; Q: 116/152, 142/187–188, 148/196.

A *rational agent* who possesses language, by contrast, can reason inferentially from premises, derive conclusions on the basis of taking some considerations to be reasons, and “by words reduce the consequences he finds to general Rules,”<sup>15</sup> i.e., derive and fully understand general precepts as reasons. Rational agents can reflectively understand, respond to, and be guided by normative reasons – they can *answer* to reasons. For Hobbes, only such creatures could have reasons, and hence be *normatively* responsible for their mental states and actions; a fundamental point of normative-reasons talk – of saying agents have reasons – is to specify when it is appropriate to counsel and to criticize them for their beliefs, passions, or actions. True, some mental states and actions are not reason-responsive: sensory perceptions, unguided mental discourse, and vital motions such as breathing are examples.<sup>16</sup> Normative responsibility accrues only to those attitudes and actions that can be reflectively guided by reasons. But insofar as rational agents’ attitudes and actions can answer to reasons, others are justified in presuming that their actions and the attitudes they express reflect their judgement of the balance of reasons, and so are warranted in imputing or *attributing* those actions and attitudes to them. Hobbes’s term of art for an agent with normative practical reasons is an “Author,” “Owner,” or “*Dominus*” of his own actions – where to have dominion over actions is not, as with the scholastics, to possess free will, but to be normatively responsible in virtue of being reason-responsive.<sup>17</sup>

One may therefore appropriately *take* rational agents to be normatively responsible for their attitudes and actions and blame them in a normative, critical sense for failing to heed their normative reasons. One might even *hold* them normatively responsible by addressing counsel or critical blame to them (such as communicating disappointment) and imposing sanctions (such as cutting off relations).<sup>18</sup> Normative blame is what Hobbes – under pressure from Bramhall – labelled “Moral dispraise” or blame, in contrast to “blame” in the more generic, non-normative sense.<sup>19</sup> It is critical, normative blame that Hobbes himself expressed at the beginning of *Elements* when he proclaimed his readers’ responsibility for weighing reasons, and characterized failures of reasoning as a fault: “if reasoning aright I winne not Consent (which may very easily happen) from them that being confident of their owne knowledge weigh not what is said, the fault is

<sup>15</sup> L 5.6: 68.

<sup>16</sup> See L 11.2–4: 22–24; 3.3: 38–40 and 8.10: 108; and 6.1: 78, respectively. Cf. Smith (2005).

<sup>17</sup> L 16.4: 244; cf. Q: 25/34. <sup>18</sup> Shoemaker (2011). <sup>19</sup> Q: 146/193.

not mine but theirs. For as it is my part [i.e., responsibility] to show my reasons, so it is theirs to bring attention.”<sup>20</sup>

In short, non-rational beings are merely causally responsible; they are appropriately blamed only in a non-normative sense, and – if they are voluntary agents – appropriately sanctioned in ways directed to their prudential capacities. But rational agents are also *attributively* responsible; they are appropriately blamed in ways that, while perhaps involving sanctions directed to their prudential capacities, also involve criticism directed to their rational capacities. Because precepts can be normative only for those who can answer to them, only rational agents can violate precepts that are normative for them – or, to use Hobbes’s vocabulary, only rational agents can *sin*, or be *guilty* or at *fault*.

We encounter Hobbes’s commitment to two dimensions of normativity – and to two corresponding kinds of normative responsibility and blame – as soon as we turn to his analysis of sin. In his first extended discussion of the concept, in chapter 14 of *De Cive*, Hobbes argued that ‘peccatum’ can be used in two senses. In the loose sense, sin encompasses all outward actions and internal dispositions (*conabitur*) “against right reason.” (This corresponds to the traditional definition in classical natural law.<sup>21</sup>) A rational agent may act or be disposed *contra rectam rationem* insofar as he could have realized, if he had reasoned rightly (*rectè*), that he is mistaken about the “meanes to the end he propounds to himselfe” and thereby acts or is disposed against his own end (*contra finem proprium*). ‘Peccatum’ in this loose sense comprises every “imprudent” action (*omne factum imprudens*) or disposition; Hobbes’s example, drawn from Matthew 7:26, is building one’s house “upon the Sand.”<sup>22</sup> Such “sins of imprudence [*peccata imprudentiae*]” comprise failures to (be disposed to) take the relevant means to one’s normative ends – whose value is ultimately grounded in one’s felicity.<sup>23</sup> In calling these prudential failures *sins*, Hobbes was invoking the *normativity* of the Prudential Precepts for rational agents, who are thereby a proper object of criticism for failing to act prudently. Non-rational creatures, by contrast, cannot sin even in this broad sense: one can judge them to have acted imprudently and so “blame” them in a non-normative sense, but precepts can be normative only for rational agents.

Yet although in acting imprudently rational agents act contrary to reason, they are not *ipso facto accountable* to others for having sinned in this loose sense. That one is *capable* of answering to reasons does not imply

<sup>20</sup> EL 1.3.    <sup>21</sup> ST I-II 7.2.    <sup>22</sup> DCv 14.16.    <sup>23</sup> DCv 14.19.

one is obligated to answer or even listen to others: mere critics or counselors have no claim against those whom they criticize or counsel. It is only in a second, “more strict” sense that ‘peccatum’ denotes actions against reasons for which one is accountable to others – actions “blamed with reason” (*ratione culpetur*) by others and therefore justifiably “call’d *malum culpae*, the evill of fault” by them.<sup>24</sup>

The loose and strict senses of sin correspond to two distinct types of normative blame or guilt (*culpa*). If *critically* blaming others consists in taking them to be responsible in the *attribution* sense that they are the appropriate object of criticism for violations, then *reactively* blaming others consists in taking them to be responsible for violations in the further, *accountability* sense that they owe one a normative response – such as a justification, excuse, or acknowledgement of wrong and hence apology, compensation, or redress – and are therefore the appropriate target of demands for such a response. If the looser sense of sin and responsibility comprises sins of *imprudence*, Hobbes identified the stricter sense with ‘injustice’, ‘injury’, and ‘wrong’.<sup>25</sup> Hobbes’s label for the second, reactive form of censure was not ‘blame’, but the juridical terms ‘accusation’ and ‘condemnation’.

The striking fact about the Hobbesian laws of nature is that intrinsically they articulate normative reasons only in the first sense at stake here: to act or be disposed in violation of natural law is intrinsically to commit only a “sin of imprudence.” It is to act or be disposed contrary to the reasons articulated by the Prudential Precepts, i.e., to fail to (be disposed to) take the relevant means to one’s preservation or felicity. Of course if one harms others in violating natural law, they will be disposed to blame one and will perhaps even seek revenge. But they would be mistaken to think one owes them any normative satisfaction: to violate natural law is not itself to fail to do anything for which one is accountable to anyone. Hobbes was explicit about this in *Leviathan*: the laws of nature themselves provide “no place for Accusation; every man being his own Judge, and accused onely by his own Conscience, and cleared by the Uprightnesse of his own Intention.”<sup>26</sup> He made this point in the context of reiterating his distinction between two types of sin. “*Sinne*” in the broader sense consists in “not onely a Transgression of a [natural or civil] Law,” i.e., “not onely in the *Commission* of a Fact [=action], or in the Speaking of Words by the

<sup>24</sup> DCv 14.17.    <sup>25</sup> DCv 14.19; cf. 3.3–5.

<sup>26</sup> L 27.3: 454. “For in the condition of Nature, where every man is Judge, there is no place for Accusation.” L 14.30: 214; cf. 30.30: 552: “there being no Court of Naturall justice, but in the Conscience onely.” Cf. Olsthoorn (2015: 24–26).

Lawes forbidden.” It also comprises the unexpressed “*Intention*, or purpose to transgresse.”<sup>27</sup> Sin in the second, narrower sense, by contrast, encompasses only outward actions that violate conventionally acquired obligations and hence civil laws. It excludes not only unexpressed dispositions to act contrary to natural or civil laws, but also outward actions that violate merely natural laws. This is because, while no one may “accuse” others for violating mere laws of nature, sin in the narrower sense – which Hobbes here called “Crime” – is “that sinne onely, whereof one man may accuse another,” that is, for which others may justly hold one accountable before a “humane Judge.”<sup>28</sup> To “accuse” and “condemn” carries heavy juridical connotations: it is not just to blame in the sense of criticizing or disapproving; it is to express normative blame *reactively* in a way that intrinsically addresses the other in seeking to vindicate one’s claims and in demanding a normative response.<sup>29</sup>

These passages from *De Cive* and *Leviathan* furnish rather explicit treatments of the two dimensions of normativity underwriting Hobbes’s ethics. The first involves reasons of the good for which one is not accountable to others, but to which one is nevertheless capable of answering and that determine whether one’s intentional states or actions are sins of imprudence. Such reasons determine whether one is the proper object of normative critical blame. The second dimension involves reasons of the right for which one is accountable to others, and which therefore determine whether one’s actions are sins of injustice and justifiably “accused” and “condemned,” and so blamed in the reactive sense that intrinsically demands a normative response accounting for one’s actions.

Although Hobbes’s distinction between the two dimensions of normativity did not receive this kind of explicit exposition in *Elements*, it nevertheless fundamentally shaped his normative philosophy even at that early stage. For it is instantiated in his well-known distinction between counsel and command. Hobbes characterized giving counsel as a form of epistemic address: it purports to *inform* others of the normative reasons they have, which obtain independently of the counsel. It presupposes counsellors’ epistemic competence; it also presupposes the counselled can answer to reasons and hence are the proper object of criticism or advice. But it does not presuppose any practical authority: counsellors have no claim-right to having their advice taken or to a normative response.

<sup>27</sup> L 27.1: 452. By “Intention” Hobbes here meant “the resolving to put some Act in execution, that tendeth thereto.” In other contexts, he used the terms “intention” or “Inclination” for a desire appearing in the course of deliberation that does not issue in action (EL 12.8; LN §27; L 6.53; 92).

<sup>28</sup> L 27.2: 452–454. See D: 42.     <sup>29</sup> D 42; LL Appendix 2.36: 1205. Cf. Wallace (1996: 79–81).

Counsellors do not even have a claim to be listened to: no one can “pretend a right to be of another mans Counsell.” Command, by contrast, appeals to a purported claim-right to be obeyed and the correlative obligation to obey, and hence a power-right to impose further obligations: it is a practical address purporting to *provide*, in virtue of the command itself, addressees with a reason for action. A “man may be obliged to do what he is Commanded; as when he hath covenanted to obey: But he cannot be obliged to do as he is Counsell’d.”<sup>30</sup>

A Hobbesian command is thus a paradigmatic example of second-personal address. The second-personal standpoint is, as Stephen Darwall has put it, “the perspective that you and I take when we make and acknowledge claims on one another’s conduct and will.” Second-personal address is addressed *to* someone in a way that intrinsically demands a normative response. By contrast, third-personal discourse is merely *about* something and does not necessarily demand a response; it may be intended to influence another’s conduct, but makes no claim to direct it. Darwall’s example is helpful. Trying to get someone to stop hurting you by informing her of the pain her action is causing involves a third-personal, epistemic address alerting her to a reason she has to stop. But *demanding* she stop is to address her directly in a second-personal way that draws on your claim against her not to treat you that way.<sup>31</sup>

If Hobbesian commands purport to *provide* addressees (new) reasons for action, demands *press* purportedly pre-existing reasons against addressees. Both presuppose a claim against addressees: the claim to be obeyed in general, and to be heeded in relation to a specific set of potential actions, respectively. If some people have normative standing to make demands on me, then I am accountable to them for conforming to the reasons they provide or press against me.

By contrast, if I fail to heed reasons I have but for which I am not accountable, then no one, not even one who has counselled me, has the standing to express reactive blame or “accuse,” “condemn,” or “punish” me in ways demanding that I account for my failure. Hobbes linked this to a further feature of the counsel-command distinction: counsellors, but not commanders, inherently adopt the pretence of serving the good of their addressees. The point is not that counsellors always *are* motivated by their addressees’ good, but that in adopting the stance of counsellor they imply they are: “he that giveth Counsell, pretendeth onely (whatsoever he intendeth) the good of him, to whom he giveth it.”<sup>32</sup> This implies that no one is

<sup>30</sup> L 25.2–5; 398–400; EL 13.5–6; 17.8; 18.8; 29.4; DCv 14.1.

<sup>31</sup> Darwall (2009: 3–8).

<sup>32</sup> L 25.3: 398.

ever in the position to “pretend a right to be of another mans Counsell; because . . . to demand right to Counsell another, argues a will . . . to gain some other Good to himselfe.” To demand a right to counsel, and thereby implicitly “to pretend benefit by it to himselfe,” is inherently incompatible with the pretence of providing counsel in the first place.<sup>33</sup> By contrast, commanders purport not only a claim-right to be obeyed once their commands have been received, but also to having their commands received in the first place.

## 5.2 Directed Obligations

The two types of normative reasons are also instantiated in the two types of *obligation* we find in Hobbes: the eudaimonistic notion he retained from classical natural-law theory and the new, juridical notion at stake in his official sense of the term. Hobbes’s strict, official notion is of a *directed* obligation that provides not just a reason for action, but one *owed* to someone to whom one is accountable. Such obligations arise, according to Hobbes, when one “lay[s] downe” or “devest[s]” oneself of a right, either by “renouncing” it simply, or by “transferring” it to others.<sup>34</sup>

In all three of Hobbes’s political treatises, the right in question, which one lays down, is in the first place a liberty-right. There is a slight evolution, however, in Hobbes’s formulation. In *Elements* and *De Cive*, Hobbes held that simply to renounce (or “relinquish”) one’s liberty-right is “by sufficient signes to declare, that it is his will no more to doe that action,” such that one thereby takes on an obligation to forebear. To transfer a liberty-right to someone specific, by contrast, is “by sufficient signes to declare, to that other accepting thereof, that it is his will, not to resist or hinder him” in the exercise of the other’s own liberty-right to act.<sup>35</sup> In the latter case, the fact that I *transfer to* someone a right “to make benefit” of his rights “without molestation” makes the directedness of my correlative obligation clear enough. In the former case, the fact that I take on the obligation via communicating to others using “signs” also intimates that the obligation is directed, but Hobbes’s early formulation does not make this explicit. *Leviathan* clarifies the issue: when a person renounces a liberty-right, “he cares not to whom the benefit thereof redoundeth.” So the benefit thereof redounds, but the person who renounces does not restrict the beneficiaries of his renunciation: the benefit redounds

<sup>33</sup> L 25.4: 400.    <sup>34</sup> L 14.6: 200.    <sup>35</sup> EL 15.3; see DCv 2.4.

unrestrictedly, to everyone.<sup>36</sup> The “benefit” in question consists in acquiring a claim-right: like the notion of *beneficium*, which ‘benefit’ connoted in this period, benefit in this context is a kind of *dominium* in a claim-right against the party who owes the correlative obligation.<sup>37</sup> If the benefit of renunciation redounds unrestrictedly to everyone, then, by contrast, when a person transfers his liberty-right, “he intended the benefit thereof to some certain person, or persons” who acquire the benefit and are owed the correlative obligation. The directedness of the obligation in both cases is explicitly confirmed in *Leviathan*’s official definition: “when a man hath in either manner abandoned, or granted away his Right; then is he said to be OBLIGED, or BOUND, not to hinder those, to whom such Right is granted, or abandoned, from the benefit of it.”<sup>38</sup>

This definition reflects two aspects of Hobbes’s account. First, an obligation in Hobbes’s strict, juridical sense is always directed, owed “to whom such Right is granted” and whom Hobbes called the “Obliger.”<sup>39</sup> Second, all directed obligations arise via convention (by which one renounces or transfers a liberty-right). Hobbes had already made the former point explicitly in *De Cive*: “no man can be obliged except it be to another.” *Leviathan* is explicit on the latter point: “there being no Obligation on any man, which ariseth not from some Act of his own; for all men equally, are by Nature Free.”<sup>40</sup> There are no pre-conventional or natural directed obligations or duties. This latter point is also implied by Hobbes’s assertion of an unconditional “naturall Right of every man to every thing,” which right “consisteth in the liberty to do, or to forbear,” and marks the absence of “Obligation.”<sup>41</sup> Hobbes had made the point in *De Cive* as well, but the earlier formulation – “obligatio enim ex pacto oritur”<sup>42</sup> – is less precise, because although a *pacto* or covenant is perhaps the most important way of laying down a liberty-right, it is not the only

<sup>36</sup> L 14.7: 200. Hobbes’s translation of this passage in Latin – “Simpliciter Renuntiat, qui medium abjicit, nemini tribuens” (LL 14: 201) – might be taken to suggest the right redounds to everyone collectively rather than individually. (It might be translated as: “Simply renouncing, which casts [it] away to people [generally], granting it to none.”) This collective interpretation, however, does not square with Hobbes’s view that a mere multitude with no collective personality is incapable of bearing rights. The more plausible interpretation is that *nemini tribuens* means “granting it to none exclusively.”

<sup>37</sup> ‘Benefit’ in this narrow (normative) sense draws on the related (legal) meaning of ‘benefice’ and ‘beneficium’ in early-modern ecclesiastical contexts. Hobbes used ‘benefit’ in this narrow sense when discussing obligation and laying down rights (e.g. EL 14.12; 15.3–8; L 11.7: 154; 14.7–8: 202–204; 15.16: 230). He used the broader sense to cover whatever contributes to a person’s good (e.g. EL 5.1; 5.13; L 5.20: 74; 12.32: 186; 13.1: 188; cf. ‘beneficium’ in DCv 2.14; 3.8; 6.19; 8.3; 9.8; 10.6–7; 13.2; 13.11; 14.12; 15.15; 17.9; 17.25). On Hobbes’s notion of property in rights, see Abizadeh (2016).

<sup>38</sup> L 14.7: 200. <sup>39</sup> EL 15.12; DCv 2.13; L 11.7: 154. <sup>40</sup> DCv 12.4; L 21.10: 336.

<sup>41</sup> L 14.3–4: 198. <sup>42</sup> DCv 8.3.

way. In all three political treatises, not only did Hobbes distinguish renouncing from transferring, he also contrasted a unilateral transfer of right, which he called “gift” or “donation,” with a reciprocal or mutual transfer of right, which he called “contract.” He further distinguished between contracts in which all parties immediately perform their part, and those in which at least one party transfers a right but is “trusted” to perform at some future time. In the latter case, the trusted party’s transfer of right consists in a “promise,” and a promise within a contract is called “covenant” or “pact.”<sup>43</sup> Hence obligation for Hobbes can be acquired via renunciation, donation, or contract – whether the contract involves covenanting or not.

Even Hobbes’s mature formulation, however, invites three possible misunderstandings. First, despite *Leviathan’s* official characterization of obligation as directed, and despite his explicit assertions that obligation so-defined arises only via convention, he nevertheless also characterized the laws of nature as intrinsically “obliging” and rational agents as intrinsically “obliged” by them, prior to any convention.<sup>44</sup> But that is because Hobbes also used ‘obliging’ and ‘obliged’ in a way that harks back to classical natural law and deviates from his own strict, officially defined sense.<sup>45</sup> Once we read Hobbes against the classical natural-law background, and discern his commitment to two dimensions of normativity, there is a straightforward explanation for the deviant uses: alongside Hobbes’s new, juridical notion of obligation, he continued to use the term in the older, eudaimonistic sense to mean the relevant means to felicity, i.e., a reason of the good. This sense of obligation – the sense in which natural laws oblige – is what Hobbes in *Leviathan* called a natural obligation: “the Law of Nature, that forbiddeth the violation of Faith,” intrinsically imposes only a “*naturall obligation*.”<sup>46</sup> This eudaimonistic obligation, retained from classical natural-law theory, is not obligation in Hobbes’s

<sup>43</sup> EL 15.7–9; DCv 2.8–9; L: 14.8–12: 202–204.

<sup>44</sup> L 15.1: 220; 15.36: 240; 19.11: 296; 30.1: 520; 40.10: 750.

<sup>45</sup> Barry (1972); Murphy (1994). Cf. Raphael (1977: 32–33).

<sup>46</sup> L 30.4: 522. Hobbes also distinguished conventional from natural obligations in DCv 15.7, but I do not rely on that passage because, given that Hobbes excised it from *Leviathan*, its significance and meaning are contentious. Hobbes there distinguished between being naturally obliged by corporeal impediments (or the laws of creation) versus by fear or hope. If the second species of natural obligation involves the fears or hopes that *actually* move us, then it differs from the sense of ‘natural obligation’ in *Leviathan*. But Hobbes suggested this second species comprises the fears or hopes that *reason dictates* to agents (*dictante scilicet ratione*) in light of their own power (*potentiori*) or weakness (*imbecillitatis propriae*) to realize their ends – in which case it is consistent with *Leviathan*. Cf. Oakeshott (1946: lix); Murphy (1994: 290–291).

strict, official sense of a conventionally incurred, directed obligation owed to others.

Second, Hobbes wrote that “he that renounceth, or passeth away his Right, giveth not to any other man a Right which he had not before,” seemingly suggesting that the obliger does not acquire any new right over the obliged, which would in turn suggest (contrary to Hobbes’s view) that the acquired obligation remains non-directed. But Hobbes was using the word “Right” strictly here, in his own technical sense to mean a liberty-right, i.e., the absence of obligation. The obliger does not acquire such a liberty-right from the transfer “because there is nothing to which every man had not Right by Nature.” But laying down one’s liberty-right does give the obliger a “benefit,” namely, ownership of a claim-right to non-resistance when the obliger exercises his corresponding liberty-right: “To *lay down* a mans *Right* to any thing, is to *devest* himselfe of the *Liberty*, of hindring another of the benefit of his own Right to the same.”<sup>47</sup> This is why when the obliger *forgives* the obligation, he gives up his acquired claim-right over the obliged, effectively enacting “a retransferring of that Right, in which the obligation consisted.”<sup>48</sup> Here we have Hobbes using the word ‘Right’ in a way that deviates, like some of his uses of ‘obligation’, from his official definition: the “Right” that in this context “consisted” in (i.e., is correlative to) an “obligation” is a claim-right, not a liberty-right. Indeed, when one has already acquired a benefit to some right from an obliged party, transferring *that* right to a third party involves not only giving up one’s original liberty-right, but also transferring the benefit of the acquired claim-right correlative to the obliged party’s obligations.

A third misunderstanding may arise from Hobbes’s insistence that to lay down a right is to obligate oneself not to “hinder” the obliger’s benefit, which suggests that obligation can only be *negative* and correlate only to a negative claim-right to non-interference.<sup>49</sup> But as Hobbes’s example of the servant who promises his master to pay a third party shows, and as his view that a subject is obligated to obey his sovereign’s commands to *do* things confirms, he had positive obligations in view as well.<sup>50</sup> Hobbes’s point in claiming that the obliged “only standeth out of his [the obliger’s] way” is once again to emphasize that, since by nature all have a liberty-right to do anything, a transfer of rights does not bestow any new liberty-rights to the obliger; but his point was neither to deny nor to characterize the

<sup>47</sup> L 14.6: 200. Cf. the definition of right in L 14.1: 198.    <sup>48</sup> L 14.26: 212.    <sup>49</sup> Green (2015: 26).

<sup>50</sup> EL 16.3; DCv 3.4; L 15.12: 228.

nature of the claim-rights (and, as we shall momentarily see, the power- and immunity-rights) gained via transfer.<sup>51</sup>

### 5.3 The Interest and the Will Theories of Direction

According to the interest theory of direction, an actor A owes the object O an obligation to undertake some action  $\phi$  if and only if O is an interested party P who stands to gain from A's  $\phi$ -ing. If Hobbes did indeed distinguish between the attributability and accountability dimensions of normativity, and so between reasons of the good and the right, then we should expect him to reject the interest theory of direction. And so he did: he conceptually distinguished between the object O of an obligation, namely the "Obliger," and an interested party P, who stands to be "dammaged" from the obligation's violation or to gain from its fulfilment; he acknowledged, moreover, that the two concepts are not coextensive. In Hobbes's technical vocabulary, when one violates an obligation, one may "damage" an interested party, but one "injures" only the obliger. To damage others is adversely to affect their interests, but to injure them is to "wrong" or commit an "injustice" against them. If one promises one's master to give money to a third party, breaking the promise wrongs the master, but merely damages the interested third party.<sup>52</sup> Moreover, it is the obliger (rather than the interested party) who "benefits" from the obligation: a benefit in this sense is a kind of ownership of a claim-right against the obliged party. (Hence to avoid confusion I refer to the interested party P as standing to gain, rather than benefit, from the obligation.)

Hobbes's distinction between the obliger and an interested party indicates he would have been alive to a standard complaint about the interest theory: that it fails to account for (potentially numerous) third parties who stand to gain from an actor's action.<sup>53</sup> But Hobbes's account also helps articulate a deeper complaint, which we can formulate as a dilemma confronting the interest theory. Either (a) "standing to gain from" is a property coextensive with "being owed" and is perhaps even what *makes* it the case that one is owed, but is nevertheless a distinct property; or (b) the two are just the same property. If the properties are distinct, then the interest theory fails to specify what it is for an obligation to be owed to another. If the properties are identical, then the interest theory fails to account for the second-personal character of *owing* something to someone.

<sup>51</sup> L 14.6: 200.    <sup>52</sup> EL 16.3; DCv 3.4; L 15.12: 228.

<sup>53</sup> For the complaint, see Sreenivasan (2010). For a response, see May (2012).

That others stand to gain from my actions is a fact fully describable from the third-personal perspective, but that I owe something to others gives them special standing to press their claim against me in second-personal address: it gives them a claim to my response.<sup>54</sup>

Since Hobbes rejected the interest theory of direction, and since he emphasized the voluntary basis of all directed obligations, it might seem that he espoused a will theory of direction, according to which A owes O an obligation to  $\phi$  if and only if O has some appropriate measure of control over A's obligation.<sup>55</sup> On a basic version of this theory, the control in question is the power-right to release A, prior to A's performance, of the obligation to  $\phi$ , along with an immunity-right against A being released, without O's consent and prior to A's performance, of the obligation to  $\phi$ . And Hobbes did think that the obliger typically has the power-right to release the one obliged to him from the obligation and an immunity-right against the obligation being released prior to performance without his consent.<sup>56</sup>

But there are at least three complications here. First, a standard objection to the will theory is that it cannot account for obligations owed to persons incapable of exercising any power-rights; Hobbes's account avoids this objection, but only by conceding that sometimes one may be owed an obligation without oneself having a power-right to release it (and without having an immunity-right against it being released). Since directed obligations can be acquired only via convention, Hobbes believed they can be owed only to persons: either to natural persons or to artificial persons who are ultimately represented by natural persons. Unless natural non-persons – such as “Children, Fooles, and Mad-men that have no use of Reason” – are artificially “Personated by Guardians, or Curators,”<sup>57</sup> they cannot acquire claim-rights and consequently be owed obligations. If natural non-persons are owed an obligation, it is because they have an authorized “representer” acting in their name, and it is the representer who possesses the power-right to release the obligation owed. Hence on Hobbes's account, the obliger's appropriate measure of control over the obligation must allow for virtual control, that is, control not truly by the obliger, but by the obliger's representer.

The second complication arises because, for Hobbes, when an obligation is owed to several persons, no particular obliger normally has the

<sup>54</sup> See Thompson's (2004: 348) analogous criticisms of Raz. <sup>55</sup> Sreenivasan (2010: 482).

<sup>56</sup> DCv 3.4; 6.14; L 15.12–13; 228; 26.6: 416. That there is an immunity-right emerges from the fact that “Men are freed of their Covenants [in only] two ways; by Performing; or by being Forgiven.” L 14.26: 212. See EL 15.12; 21.12; DCv 2.15.

<sup>57</sup> L 16.10: 248.

power-right unilaterally to release the obligation. In the normal case, a particular obliger has the power-right to release another only from owing the obligation to himself. The power-right is over the obligation's particular "rays" of directionality, but the obligation is extinguished only when all of its directional rays are. This feature comes out nicely in Hobbes's account of the political obligation to obey the sovereign. According to all three of his political treatises, political obligation originates, in the case of an "instituted" commonwealth, in a series of covenants in which each would-be subject individually promises to each other one to "give up" to a common sovereign the "Right of Governing" himself if the other does so as well.<sup>58</sup> According to *Leviathan's* mature formulation, each subject covenants also to "Authorise all of his [the sovereign's] actions" such that the sovereign shall "represent" and "beare their Person."<sup>59</sup> Since the one who will be instituted as sovereign is not himself party to any of these contracts,<sup>60</sup> the subsequent transfer of the right of governing, which each had promised to undertake, consists in a donation to him.<sup>61</sup> The resulting obligation to obey the sovereign is therefore owed to each other subject (who covenants) as well as to the sovereign (who accepts the donation).<sup>62</sup> That it is owed to each other subject explains why no mere fraction of subjects can release anyone from his political obligation: "for they are bound, every man to every man," as Hobbes put it in *Leviathan*, "so that any one man dissenting," he would still be bound even if all the rest should agree to release him. That political obligation is owed to the sovereign as well – since "they have also every man given [i.e., donated] the Sovereignty to him" – explains why, even if all the subjects were to agree, they still could not "without his permission" release themselves of their political obligations.<sup>63</sup> The contrast between private contractual obligations, which are directly owed only to fellow subjects, and political obligation, also owed directly to the sovereign, is instructive: in a commonwealth, "private men may remit to one another their debts; but not the robberies or other violences, whereby they are endammaged; because the detaining of

<sup>58</sup> L 17.13: 260. Cf. EL 19.7; 21.2; DCv 5 and 6.

<sup>59</sup> L 17.13: 260. Authorization (or commission) is distinct from the transfer of right. D: 54–55. Cf. Copp (1980); Green (2015).

<sup>60</sup> DCv 6.20; L 18.4: 266; 21.II: 336.

<sup>61</sup> The transfer of right to the sovereign is a "iuris donatione." DCv 6.20. Warrender's (1957: 135–136) argument ignores this donation, and Gauthier's (1969: 104) explanation for *Leviathan's* new language of authorization rests on the mistaken assertion that subjects "make no deed of gift" to the sovereign in *De Cive*.

<sup>62</sup> DCv 6.20; L 18.3: 264. *De Cive* refers to this as the "duplici obligatione civium."

<sup>63</sup> L 18.3: 264.

Debt, is an Injury to themselves; but Robbery and Violence, are Injuries to the Person of the Common-wealth” as well.<sup>64</sup> So political obligations are owed to fellow subjects, even though subjects neither individually nor jointly have the power-right to release such obligations. This is because being owed an obligation is not essentially a matter of having the normative power to release; it is, rather, a matter of having the normative standing to reiterate and vindicate it in a way that demands a normative response from the obliged.

The case of political obligation also illustrates the third complication in Hobbes’s account, which arises from the fact that sometimes an obliger has the power-right to release the obliged from owing the obligation to *others* as well. This exceptional case may arise when an obligation is owed to others only on condition that it *also* be owed to some obliger in particular – whom we may call the *essential obliger*.<sup>65</sup> Thus the Hobbesian sovereign has the power-right unilaterally to release his subjects from the obligation to obey; he may even “sell, or give his Right of governing to a stranger.”<sup>66</sup> When the sovereign either renounces or transfers his right to rule, he thereby releases his subjects from the obligation to obey him, which they owe not only to himself, but also to each other. He can do this in part because he is his subjects’ representer. But representation is insufficient fully to explain the relationship between the obligation as owed to the sovereign and as owed to subjects. This insufficiency is illustrated by the case of the sovereign who *unwillingly* loses the capacity to maintain peace and protect his subjects. According to Hobbes, this loss of control releases subjects from their obligation to obey him<sup>67</sup> – not just as it is owed to the sovereign but also as owed to fellow subjects. This is because political obligation is owed to subjects *conditionally* upon it *also* being owed to the sovereign, the essential obliger. (This is why if a sovereign does not accept the donation of sovereign right, the attempted transfer fails, and would-be subjects do not owe it to each other to obey him.) A sovereign not only has the claim-right to be obeyed and power-rights to impose new obligations and to give up his claim-right, but also a further power-right over and immunity-right against his subjects: he can cancel their political obligations to each other, but they cannot justly remove his ownership of the right to rule without his consent. Although some such immunity-right is normally joined to being owed an obligation, the sovereign once again illustrates the exceptional case: where an obligation is owed to an obliger conditionally upon it also

<sup>64</sup> L 15.12: 228. Cf. DCv 3.4.      <sup>65</sup> Cf. Grotius JBP II.21.11.

<sup>66</sup> L 19.23: 304. Cf. EL 21.9; DCv 9.13.      <sup>67</sup> L 21.21: 344.

being owed to an essential obliger, the former will not have an immunity-right against the essential obliger over the release of the obligation. The sovereign illustrates this normative relationship in cases beyond the obligation to obey as well. According to Hobbes, subjects lack an immunity-right against the sovereign over a broad range of claims: it is within the sovereign's right to tax, expropriate, or redistribute their property as he sees fit.<sup>68</sup>

#### 5.4 Owing an Obligation versus Being Held Accountable

These complications demonstrate that Hobbes's theory distinguishes two properties that simple will theories of direction conflate. Although Hobbes believed that being owed an obligation normally goes in tandem with having some measure of control over it, he also recognized that the two properties are not always coextensive. On Hobbes's account, 'being owed an obligation' neither refers to the same property as, nor means the same thing as, nor is made the case by having the power-right to release the obligation. Rather, being owed an obligation reduces to having the standing to hold the obliged accountable, akin to having a Grotian "perfect right."<sup>69</sup>

Nor is it the case for Hobbes that the genuinely normative concept of *having an obligation* or *being accountable* is reducible to a non-normative, naturalistic concept such as *being motivated* or *being held accountable* – as John Selden, who might be characterized as an ethical-naturalist reductionist, took them to be.<sup>70</sup> Hobbes was explicit about this, as in a 1647 addition to *De Cive*:

*To be obliged* [esse obligari], and *to be tyed being obliged* [obligatum teneri], seems to some men to be one, and the same thing . . . [but] I say thus, That a man is "obliged" by his contracts, that is, that he ought to performe for his promise sake; but that the [punishment appointed by the] Law "tyes him being obliged," that is to say, it compells him to make good his promise, for fear of the punishment appointed by the Law.<sup>71</sup>

Hobbes's assumption that directed obligation is not reducible to motivation or fear of punishment is also manifest in – indeed, essential to – his

<sup>68</sup> EL 24.2; 27.8; DCv 6.15; 8.5; L 24.7: 390; 29.10: 504. <sup>69</sup> JBP I.1.5.

<sup>70</sup> The view that Hobbes, like Selden, reduced obligation to punishment and, ultimately, to a kind of motivation is advanced by Tuck (1979: 91–92, 126–127). See Darwall (1995); Cromartie (2011). For a defence of an ethical-naturalist view reducing accountability to practices of holding accountable, see Watson (2004: chapter 8).

<sup>71</sup> DCv 14.2, italics reversed.

treatment of oaths. Hobbes argued in all three political treatises that swearing an oath by which one calls on divine punishment for violating a covenant may help to strengthen one's *motivation*, but it "addeth not a greater obligation to perform the Covenant sworne, than the Covenant carrieth in it self."<sup>72</sup> No reading of Hobbes as a reductionist can adequately explain his refusal to reduce obligation to motivation, or reduce being accountable to being held accountable.

Although it is not possible to *reduce* Hobbes's concept of being accountable to the non-normative, naturalistic concept of being held accountable, the former – and the related notions of obligation, injury, and wrong – is nevertheless partly constituted by and so intrinsically dependant on the latter. Being accountable to others reduces to their (i) having the normative standing (ii) to hold one accountable. We can specify the notion of accountability undergirding Hobbes's ethics more precisely. For me to be *culpable* towards you is for me to fail to conform to a reason for which I am accountable to you. For me to *be* accountable to you for conforming to a reason is for you to be *pro tanto* justified to hold me accountable for it – i.e., for you to have the normative standing to hold me accountable.<sup>73</sup> And for you to *hold* me accountable is for you to be disposed to respond to me, if and because you believe I am culpable towards you, (1) by expressing (a) reactive blaming passions intrinsically addressed to me, or by expressing (b) the belief that such passions would be justified, (2) in ways that, (a) on the one hand, seek to reiterate and vindicate your demand that I conform to the reason for which you believe me accountable and that, (b) on the other hand, communicate an auxiliary demand for a normative response accounting for my actions. (I take *reactive attitudes* to include both (a) reactive conative mental states, or what I call *reactive passions*, and (b) cognitive beliefs or judgments that such reactive passions would be appropriate.<sup>74</sup>)

We should distinguish this notion of holding someone accountable, and the blaming responses involved, from the two other kinds of responsibility and blame outlined earlier. You might, first, take me to be *causally* responsible for a painful or unwelcome event and hence you may, besides grief or fear, experience passions directed towards me such as anger or

<sup>72</sup> EL 15.17. See DCV 2.12; L 14.31–33: 216–218.

<sup>73</sup> Smith (2007) argues one cannot analyze being accountable in terms of the appropriateness of holding accountable, because sometimes one has relationship-based reasons to avoid holding culpable friends accountable. But an analysis in terms of *pro tanto* appropriateness obviates this objection.

<sup>74</sup> On reactive attitudes, see Strawson (2008); Wallace (1996); Darwall (2009).

disappointment, or simply believe such passions are justified, or both. You may also retaliate by harming me for violating your expectations, that is, you may exact “Revenge,” which Hobbes defined as “retribution of Evil for Evil.”<sup>75</sup> Such blaming responses may be appropriate even if I were a non-rational beast: your responses are directed at me but not in any sense that presumes my capacity for answering to reasons. I am the mere object of your passions or belief, and your sanctions impose prudential incentives.

But because I am a rational creature, you might, second, also take me to be responsible in a normative, attribution sense: you might attribute my voluntary actions to me insofar as you take me to be capable of answering to reasons. Your *critical* blaming responses will therefore target this reflective capacity. They may even be addressed to me – and hence appropriately involve anticipating (though not demanding) a normative response – but this is a contingent matter: you may merely feel and express contempt for or disgust towards me.

But if, third, you take me to be *accountable* to you for the unwelcome event, then you will experience, or believe you would be justified in experiencing, passions not only directed at but intrinsically addressed to me, insofar as they at least latently *demand* a normative response from me. The accusatory and condemning attitudes involved in taking someone to be accountable are reactive, second-personal attitudes.<sup>76</sup> And they target the interpersonal, joint capacity people have to bind their will to others in virtue of the interpersonal, dialogic capacity to produce and interpret signs of their wills and hence to offer and accept these signs as binding.

Contemporary moral philosophers call the paradigmatic reactive passion resentment; Hobbes called it *revengefulness* or, as he put it in Latin, *Vindicta* or *cupido Vindictæ*.<sup>77</sup> It is true that revenge, which Hobbes understood to consist in retaliatory harm, need not itself take a reactive form: when beasts retaliate, they do so without demanding any normative response. But the passion of revengefulness as Hobbes conceived it cannot be experienced by non-rational creatures because it is not merely a desire for revenge in this generic sense. Hobbes’s first pass at a definition, in *Elements*, characterizes revengefulness as a desire not only to inflict harm and make one’s target regret his action – to make “him that hath hurt us, to find his owne action hurtfull to himselfe” – but also to

<sup>75</sup> L 15.19: 232.

<sup>76</sup> Strawson has been criticized for not adequately distinguishing between second-personal, reactive attitudes and other kinds of attitudes experienced in virtue of participating in relationships (Wallace 1996: 27–28; Darwall 2009).

<sup>77</sup> EL 9.6; L 6.34: 86; LL 6: 87; DH 12.4.

force him outwardly to “acknowledge” his regret (which “is soe difficult, that many a man had rather dye than doe it”).<sup>78</sup> This is a first pass because, although expressing regret here involves a normative response, it involves acknowledging only self-regarding reasons of the good (insofar as one’s actions have turned out to have painful consequences for oneself). Hobbes’s mature formulation in *Leviathan* characterizes the passion in more precisely reactive, second-personal terms: revengefulness here becomes “*Desire*, by doing hurt to another, to make him condemn some fact [=action] of his own.” Revengefulness demands not mere expressions of prudentially inspired regret over past actions, but *condemnation* of those actions in the face of accusation, i.e., acknowledgement of guilt for an injury in virtue of having violated a valid normative demand.<sup>79</sup> Hobbes’s formulation in *De Homine* reiterates this link to acknowledging guilt at having committed a wrong: *cupido Vindictæ* is “the constant and long-lasting will of doing evil to someone in order that he repent [*pæniteat*] of the supposed injury done by him.”<sup>80</sup> (Revengefulness was thus Hobbes’s analogue to Aristotelian *orgē* or “anger,” which the philosopher had defined as a desire for perceptible retaliation against perceived injustice.<sup>81</sup>)

One can experience reactive blaming attitudes such as revengefulness without outwardly expressing them. But whereas some contemporary moral philosophers deem private reactive blame sufficient for holding others accountable,<sup>82</sup> in the Hobbesian framework it is not. We can put this by distinguishing between *taking* others to be accountable and *holding* them accountable. For you to take me to be accountable to you is for you to be disposed to respond reactively to me in ways intrinsically addressed to me, or to believe that responding second-personally in these ways would be *pro tanto* justified, if and because you believe I am culpable towards you. And for you to respond reactively consists either in having reactive blaming attitudes towards me or in outwardly expressing such attitudes. Having unexpressed reactive attitudes is therefore sufficient for taking others to be accountable. But for Hobbes it is not sufficient for holding them accountable because doing so involves addressing them in order to demand a normative response – which one cannot do except by communicating one’s attitude.

For Hobbes, the paradigmatically second-personal blaming actions required for holding others accountable are those to which revengefulness typically gives rise: sanctioning others in ways that involve “accusing” and “condemning” them to elicit self-condemnation. What makes expressed

<sup>78</sup> EL 9.6.    <sup>79</sup> L 6.34: 86.    <sup>80</sup> DH 12.4.    <sup>81</sup> Rh 2.2.    <sup>82</sup> Wallace (1996: 56, 77).

reactive responses instances of holding another accountable is that they simultaneously do three things. First, they impose sanctions by harming or threatening to harm the other; they therefore provide *incentives* to comply with one's demands. This element is shared with the non-normative blame one expresses when holding others causally responsible, which is what Hobbes meant by the English word 'blame' in its most generic sense. Second, they *express* expectations – as is done in counsel or criticism – about the other's actions and accompanying attitudes, as well as the belief that those expectations have been violated. This element is shared with the normative, critical blame one expresses when holding others attribution-responsible, which Hobbes characterized using the normative vocabulary of sin, duty, guilt, law, etc. Finally, they communicate the *demand* for a normative response from the other: they are not merely ways of expressing one's passions and beliefs – as one might do in giving counsel or expressing sadness or disappointment – but also of seeking acknowledgment of the normativity of the violated demand. This is the reactive element unique to holding others accountable, which Hobbes characterized in terms of “accusation” and “condemnation.”

Hobbes was greatly exercised by the potentially devastating effects of the usual human practices of holding others responsible. Because non-rational creatures “have noe conception of right and wronge, but onely of pleasure and paine,” they do not engage in “censure of one another . . . as long as they are themselves at ease.”<sup>83</sup> The “censure” they do express in response to harm is not reactive or accusatory blame, because they lack the capacity to demand justificatory reasons from others; indeed, it remains a non-normative kind of blame because they lack even the capacity to understand, respond to, and be guided by reasons. But practices of censuring are transformed in four important ways amongst linguistic creatures who can answer to reasons and “distinguish between *Injury*, and *Dammage*.” First, humans are disposed to hold others *normatively* responsible for hurting them: “men are accustomed to regard as *evill* [malos putare], that is to say, to *lay some fault* to their charge [culpam attribuere] from whom they receive *evill*.”<sup>84</sup>

Second, rational creatures censure not just harms, but also injuries, even when these are harmless in themselves: “irrationall creatures . . . as long as they be at ease, they are not offended with their fellowes: whereas Man is then most troublesome, when he is most at ease.”<sup>85</sup> The reason we respond even to harmless injuries is that they force us to contemplate our

<sup>83</sup> EL 19.5.    <sup>84</sup> DCv 14.17.    <sup>85</sup> L 17.11: 260. Cf. DCv 5.5.

powerlessness relative to wrongdoers, who act as if they could shirk their obligations to us with impunity. “Iniury how light soever be the damage, is alwayes grievous, as putting us in minde of our disability to helpe ourselves. And into Envy of the power to doe us wronge.”<sup>86</sup> We respond because even harmless injuries end up harming us by inflicting mental pain or “grief.”

Third, we are disposed to respond *reactively* to injuries – experiencing revengefulness or reactive forms of hatred (*odium*) and envy (*invidia*) – because, unlike non-rational creatures, we desire second-personal “acknowledgement” or recognition from others: “men are continually in competition for Honour and Dignity, which these [irrational] creatures are not; and consequently amongst men there ariseth on that ground, Envy and Hatred, and finally Warre; but amongst these [creatures] not so.”<sup>87</sup> When we take others to have violated their obligation to us, we are prone to reiterate and vindicate our expectation by demanding a response from wrongdoers acknowledging the normative validity of our demand.

Fourth, once we conceive of right and wrong and relations of accountability, we are disposed to respond to mere *harms* reactively too.<sup>88</sup> Humans are so prickly, according to Hobbes, that we often respond with anger and hatred and seek revenge over mere “trifles, as a word, a smile, a different opinion, and any other signe of undervalue” or contempt. Of course anger, blame, and revenge need not be reactive or second-personal: one might merely inflict harm without demanding any normative response or acknowledgement. But humans are prone to specifically reactive hatred, envy, blame, and revengefulness precisely because we desire and demand others’ acknowledgement. Each demands that others “should value him, at the same rate he sets upon himselfe: And upon all signes of contempt, or undervaluing, naturally endeavours . . . to extort a greater value from his contemners, by dompage; and from others, by the example.”<sup>89</sup> The harm we inflict on others in holding them accountable for having offended us is a way of demanding they acknowledge the normative validity of our expectation to be honoured. It is because we desire and demand others’

<sup>86</sup> EL 24.2.     <sup>87</sup> L 17.7: 258. Cf. EL 19.5; DCv 5.5.

<sup>88</sup> According to *Elements*, revengefulness arises against “him that hath hurt us”; the definition in *Leviathan* does not specify what kind of “fact” provokes the passion; and although the definition in *De Homine* refers to injury, it refers more specifically to *supposita injuria*. EL 9.6; L 6.34: 86; DH 12.4.

<sup>89</sup> L 13.5–6: 190–192. See DCv 14.17: “the prosperous actions of enemies (because they increase their honours, goods, and power) and of equals, (by reason of that strife of honours which is among them) both seeme and are *irksome* [molesta], and therefore *evill* to all.” Cf. EL 14.4; L 13.5–7: 190–192; 17.7: 258.

acknowledgement that we imagine they owe it to us to fulfil our expectation not to be harmed.

Hobbes discerned great danger in this fourfold transformation. It is true that, even though we do not have the *standing* to hold others accountable for non-injurious harms, by right of nature we may nevertheless respond however we see fit. But our state-of-nature reactive responses to injury and harm risk deteriorating into forms of cruelty that imprudently aggravate conflict. Hobbes sought to counter this danger by three reforms to the usual practices of holding responsible. First, he taught his readers that others are accountable to them only for directed obligations incurred through renunciation, donation, or contract. Unless they have laid down their right of nature, humans lack any obligations to refrain from harming each other. Hobbes went so far as to say people do not have any natural directed duty to refrain from the “Deceit and Violence [*vim & dolum*]” he compared to the predatory “rapacity of wild beasts [*ferinam rapacitatem*].” In saying this, Hobbes intended to reform the “inbred custome” of hurling the term ‘rapacity’ against each other to express reactive blame or “mutual reproach” (*pro conuitio inuicem obiiciant*).<sup>90</sup> Hobbes castigated the imprudent overextension of reactive blaming responses to not only “Corporeall” pain, but also the “Phantastical” pain we feel when others inevitably disappoint our desire for acknowledgement. Absent conventions by which “the publique worth of a man” is set by the sovereign, it is unrealistic to expect others to value one as highly as one values oneself.<sup>91</sup> Such an expectation runs afoul of everyone’s natural tendency to overestimate their own worth and to rank themselves higher than most,<sup>92</sup> and will lead to imprudent dispositions and behaviour in violation of the ninth law of nature against “Pride,” which prescribes “*That every man acknowledge other for his Equal by Nature.*”<sup>93</sup>

Second, Hobbes insisted that one should not retaliate against harms or injuries – whether reactively or not – unless doing so prospectively serves one’s ongoing, future good. Here Hobbes was seizing on a longstanding motif in the history of European ethics concerning practices of revenge and punishment. Grotius, who had painstakingly documented this tradition a few decades earlier, located the motif’s origin in Plato’s contention that punishment is to be exacted not because of a past wrong – “for what is done can never be undone” – but to prevent future wrongdoing.<sup>94</sup> Appropriating Plato’s insight, Seneca had in turn linked retrospective

<sup>90</sup> DCv Epistle Dedicatory.2. The Cook translation is unreliable here.

<sup>91</sup> L 27.20: 466; 10.18: 136.    <sup>92</sup> L 13.2: 188.    <sup>93</sup> L 15.21: 234.    <sup>94</sup> *Laws* 934a; 854d; JBP II.20.4.

punishment to an Aristotelian conception of anger<sup>95</sup> (the analogue to Hobbesian revengefulness): “We are not to harm a man because he has sinned, but that he may not sin; punishment is to refer not to the past, but to the future, imposed not from anger [*irascitur*], but as precaution [*cavet*].”<sup>96</sup> According to Seneca, punishing out of a desire for revenge is never *utile* and, even if it were, would not be *honestus*.<sup>97</sup> Even worse than retrospective revenge, moreover, is harm inflicted without reference to either past or future, but merely to take pleasure in another’s suffering; this Seneca condemned as cruelty (*crudelitatem*) or inhuman bestiality (*feritas*).<sup>98</sup>

Hobbes intensified this tradition by collapsing Seneca’s distinction between retrospective revengefulness and the bestial cruelty of taking pleasure in others’ pain: he castigated *any* infliction of harm not oriented to future good as mere “*Cruelty*.” Not only is revenge for its own sake – i.e., retaliatory harm inflicted with the intention of experiencing the pleasure of satisfying a desire for revenge – cruel, so too is any retaliation lacking the *positive* aim of fostering future compliance with one’s demands. Both points, negative and positive, are combined in Hobbes’s statement of, and argument for, the seventh law of nature against cruelty, “*That in Revenges, (that is, retribution of Evil for Evil,) Men look not at the greatness of the evil past, but the greatness of the good to follow. Whereby we are forbidden to inflict punishment with any other designe, than for correction of the offender, or direction of others.*” Hobbes’s argument is instrumental: cruelty is contrary to reason because “Revenge without respect to the Example, and profit to come” engenders in avengers only the momentary pleasure, experienced while inflicting harm, of presently satisfying their desire for revenge, without any ongoing, anticipatory pleasure concerning the future. But this is merely a *present* and fleeting “triumph, or glorying in the hurt of another, tending to no end; (for the End is always somewhat to Come;) and glorying to no end, is vain-glory, and contrary to reason; and to hurt without reason, tendeth to the introduction of Warre.”<sup>99</sup> Rather than ongoing pleasure, the long-term effect of cruelty is to provoke an escalating cycle of retaliation, because cruelty expresses either hatred or contempt, and “most men choose rather to hazard their life, than not be revenged” when subjected to “signes of hatred, or contempt.” This is precisely the point of Hobbes’s eighth law of nature, “*against*

<sup>95</sup> DI 1.2.3; 2.1.1.    <sup>96</sup> DI 2.31.8; cf. 1.19.7.    <sup>97</sup> DI 1.9.1; 1.12.6; 2.32.1.    <sup>98</sup> DI 1.6.4; 2.5.1–3.

<sup>99</sup> L 15.19; 232. Cf. EL 16.10; 18.11, which notes that by natural law revenge “ought not to aime . . . at present delight. but at future profit”; DCv 3.11; 4.9; LN §14.

*Contumely*,” which prescribes “*That no man . . . declare Hatred, or Contempt of another*,”<sup>100</sup> and this prudential point applies no less to the sovereign than to private individuals.<sup>101</sup>

Hobbes also sought to immunize the traditional motif against the incursion of any intrinsically retrospective, retributivist role for retaliation. The classical natural-law theory bequeathed by Aquinas had found room for such a role – despite Aquinas having agreed with Seneca that vengeance is illicit if one’s intentional object is (taking pleasure in) another’s harm, and licit only if one’s intention is some good. For amongst the goods at which retaliation might aim Aquinas included not just reform and prevention, but also justice. And while reform and prevention are prospective goods, aimed at preventing “future sin,” Aquinas characterized seeking justice as intrinsically retrospective: it aims at “healing the past sin” by restoring the cosmic order’s balance or “equality” in counterbalancing the sinner’s will.<sup>102</sup> Bramhall tried pushing this classical view on Hobbes, arguing that even if punishing a guilty person performed no useful role, but was purely “vindicatory,” it would still be warranted insofar as it would satisfy “a zeal of Justice, by giving to every one his own.”<sup>103</sup> Hobbes replied tersely: the Bishop “taketh punishment for a kind of revenge” upon the criminal, i.e. undertaken for its own sake, rather than as an instrument used “to reforme the will of him or others.”<sup>104</sup>

Hobbes’s purely prospective emphasis on deterrence may even seem to rule out as cruel any retaliatory actions that express reactive attitudes for past wrongs – responses which I have argued are constitutive of the second dimension of normativity. Hobbes claimed “the end of punishing is not revenge, and discharge of choler; but correction, either of the offender, or of others by his example.”<sup>105</sup> But once we read Hobbes in light of the Platonic-Senecan motif he was appropriating, this impression dissipates. As Grotius put it, taking pleasure in “the bare spectacle of an enemy’s suffering, is nothing but a false and imaginary good,” and it was “in this sense that vengeance amongst men was condemned” by those, such as Seneca, who insisted on retaliation’s prospective nature.<sup>106</sup> Similarly, when Hobbes claimed “the end of punishing is not

<sup>100</sup> L 15.20: 234.

<sup>101</sup> EL 16.11 criticizes the humiliating treatment inflicted on subjects in the course of holding them judicially accountable: the contemptuous “reproaches . . . of those that sitt in place of Judicature, toward those that are accused at the barre” is a custom that “hath prevayled” despite the fact that “to greive them in that manner be no part of the punishment, for their Crime. nor containd in their office” and “contribute nothing towards” the reasonable aim of punishment.

<sup>102</sup> ST II-II.108.1; 108.4. <sup>103</sup> Quoted in Q: 123/162. <sup>104</sup> Q: 134/177. <sup>105</sup> L 30.23: 542.

<sup>106</sup> JBP II.20.5.

revenge,” he meant revenge *for its own sake*, i.e., for the pleasure of satisfying the desire of revengefulness. Whatever reactive attitudes are expressed in retaliation must be expressed not for “discharge of choler,” but ultimately for the purely *instrumental* purpose of reform or deterrence. Thus when Hobbes criticized Bramhall for taking “*blame* not for the dispraise of a thing, but for a prætext and colour of malice and revenge,”<sup>107</sup> he was not ruling out normative or even reactive forms of censure or punishment. Rather, beyond defining ‘blame’ in its most generic, non-normative sense, he was condemning retributivist views that countenance revenge for its own sake.

Hobbes’s third and most important response to the dangers of revengefulness was to argue that the practice of holding others accountable – indeed, of holding others attribution-responsible – should be placed in public hands. Unlike Seneca’s purely ethical remedy, which sought to eradicate revengefulness, Hobbes proffered a political response. His central claim was that holding others responsible as private persons will thwart the point of the practice – to fulfil our expectations. Although people customarily call whatever pains them “*malum culpæ*, the evill of fault; but yet if any thing be blamed [*culpetur*] it is not presently to be term’d a *sinne* [peccatum], or *fault* [*culpa*], but onely if it be blamed with reason [*ratione culpetur*].” Left to our own devices, “we all measure *good* and *evill* by the pleasure or paine we either feele at present, or expect hereafter.” We therefore end up disagreeing about what is evil and hence blame-worthy: “through the diversity of our affections, it happens that one counts that *good*, which another counts *evill*.” Because evaluative disagreement breeds conflict, to avoid war it is prudentially necessary to transfer the right of judging culpability and holding others responsible to the sovereign: because given “so great a diversity of assessments [*censentium*],” “its impossible to define by the consensus of single men whom the same things do not please and displeas, what actions are culpable [*culpanda*], and what not culpable . . . the commonwealth is to determine what is *blamed with reason* [*ratione culpandum*].” Hobbes wanted to reform the customary practices of holding responsible so that the objects of both critical and reactive censure would be determined by a public practice of legislation, adjudication, and punishment: “So that a *fault* [*culpa*], that is to say, a *SINNE*, is that, which a man do’s, omits, sayes, or wills, against *the reason of the commonwealth*, that is, contrary to the Lawes.”<sup>108</sup>

<sup>107</sup> Q: 39/53.      <sup>108</sup> DCv 14.17.

### 5.5 Legal Accountability and Punishment

The upshot is that, for Hobbes, holding accountable via legal punishment serves not just a deterrent and reforming function, but also an *expressive* and *dialogic* one – precisely insofar as the latter serve an instrumental, deterrent function.<sup>109</sup> If punishment is not aimed at deterrence, it is cruel; and if it is cruel, it is mere “hostility,” i.e., it is punishment in the broad sense of sanction or revenge, but not in the narrow and proper sense: “all evill which is inflicted without intention, or possibility of disposing the Delinquent, or (by his example) other men, to obey the Lawes, is not Punishment; but an act of hostility.”<sup>110</sup> But according to *Leviathan*, the lack of cruelty, albeit a necessary condition of punishment proper, is insufficient. To qualify as punishment in the strict sense distinct from hostility, the harm must be inflicted by a *public authority*: merely “private revenges,” even if aimed at deterring offenders and hence not cruel, cannot “properly be stiled Punishment; because they proceed not from publique Authority.”<sup>111</sup> Furthermore, punishment must be preceded by a legal procedure that begins with accusation and ends with public *condemnation* (finding of guilt): “the evill inflicted by publique Authority, without precedent publique condemnation, is not to be stiled by the name of Punishment; but of an hostile act; because the fact for which a man is Punished, ought first to be Judged [LL: *probari & judicari*, tried & judged] by publique Authority, to be a transgression of the Law.”<sup>112</sup>

The aim of public accusation and condemnation is not the pleasure of giving vent to feelings (the “discharge of choler”). Its function is not *merely* expressive. Its expressive function is essentially dialogic: to reiterate the initial (legal) requirement and to demand and elicit a normative response from the accused that acknowledges the requirement. What makes punishment dialogic is the requirement that it be public: punishing becomes simultaneously a way for the public authority to reiterate the violated requirement and for the *offender* to acknowledge the requirement’s validity. This is because on *Leviathan*’s theory of authorization and representation, the public authority who punishes a condemned subject does so in the name of not only the commonwealth, but also the condemned subject himself, who authors all of the sovereign’s actions as his own: in virtue of being “acts of publique Authority,” punishments also have for their

<sup>109</sup> On the expressive function of legal punishment, see Feinberg (1965). On its dialogic function, see Duff (2001).

<sup>110</sup> L 28.7: 484.    <sup>111</sup> L 28.3: 484.    <sup>112</sup> L 28.5: 484–485.

“Author, the person condemned.”<sup>113</sup> Punishment serves not only as a way for the public authority to reiterate the initial requirement by way of condemning its violation; it also serves as a way to express the offender’s *response* to that reiteration.

Although private revenges cannot themselves express the offender’s acknowledgement of the requirement’s validity, they do appropriately aim at eliciting such a response. Revenge in part performs its deterring function *thanks to* its expressive and dialogic functions: it reiterates the initial demand by giving expression to negative reactive attitudes at its violation, and it communicates the demand that the offender acknowledge the initial demand’s validity.<sup>114</sup>

Insofar as punishment is public – insofar as subjects transfer to the sovereign the right to hold people accountable to each other – not all means of holding others accountable may be *directly* available to every obliger. The sovereign may decide that private obligers have standing directly to express certain reactive passions via legal channels, by pursuing a legal suit, for example, but not to impose coercive sanctions themselves. Private parties to legal contracts each owe it to the others to fulfil the terms of contract, but none may have the liberty-right *directly* to hold the others accountable via coercion: typically, only publicly authorized agents may coercively punish such violations.

It might therefore be objected that, contrary to my characterization, ‘owing an obligation to someone’ and ‘being accountable for an obligation to someone’ are not identical: one owes the fulfilment of a private contractual obligation to the other parties, but is (supposedly) accountable only to the state (since only the state can directly hold one accountable by coercively punishing violations). But the fact that only the state can coercively punish violations of private law does not mean the private parties to the contract are accountable only to the state rather than to each other: it means the state holds the contracting parties accountable *to each other*, and this *on behalf of* the contracting parties. The parties are still accountable to each other, but they coercively hold each other accountable indirectly, via a third party acting in their name.<sup>115</sup> It is true that according to Hobbes subjects are *also* directly accountable to the sovereign, i.e., they also owe it to their sovereign to uphold their legal contracts. But this simply means that, in coercively punishing violations of private law, the state does double duty: it upholds legal contracts both directly, on its own behalf, and

<sup>113</sup> L 28.6: 484; cf. 17.13: 260; 18.1: 264; 20.1: 306; 24.7: 390. <sup>114</sup> See EL 18.11.

<sup>115</sup> For a contemporary interpretation of torts law along these lines, see Goldberg and Zipursky (2010).

indirectly, on behalf of the private parties. The fact that obligated parties are also accountable to private obligers is reflected in the fact that the latter typically have a right of *vindicatio* – the right to accuse violators in a legal procedure geared to the legal condemnation and punishment of violators.

This reinforces the point that having the standing to hold obligated persons accountable is distinct not only from having normative control over the obligation, but also from having the liberty-right to enforce it, i.e., to *cause* obligated persons to conform to their obligation by providing incentives to do so. The difference is illustrated by the contrast Hobbes drew between the sovereign who merely exercises his liberty-right in retaliating against foreign *enemies*, and the sovereign who exercises authority over his *subjects* in holding them accountable for violations of law via *punishment* properly understood.<sup>116</sup> The liberty-right to enforce via sanctions involves a third-personal standing that does not intrinsically demand any answer or acknowledgement from the other. One may have such a liberty even in relation to non-rational beasts. Only the relation of accountability involves a *second-personal, practical* standing that intrinsically demands a response from the one punished. To equate ‘x has the liberty-right to coerce y’ with ‘y is accountable to x’ is to conflate the right of nature enjoyed by *all* rational agents in the Hobbesian state of nature with the second-personal practical standing that emerges only once persons undertake directed obligations.

In sum, Hobbes’s account requires distinguishing between the following relations to obligation: (1) standing to gain from its fulfilment (and to be harmed or “dammaged” by its violation), which presupposes an *interest* in its fulfilment; (2) having the standing to criticize, commend, and advise others in its light, which presupposes the epistemic *competence* to apply the corresponding reason (but not any claim to a hearing or response); (3) having the standing to release it or to prevent its release, which involves the appropriate measure of normative *control* (power- and immunity-rights); (4) having the standing to cause an obligated person to conform to the obligation, including a *liberty-right* to enforce it via sanctions; and (5) having the standing to hold the obligated person accountable, which is a practical *authority* over the obligated person. The first relation involves a third-personal, merely *pragmatic* standing; the second involves a third-personal theoretical or *epistemic* standing; and the third and fourth involve a third-personal, albeit *practical* standing. It is the fifth

<sup>116</sup> L 28.13: 486.

and final relation of accountability – a genuinely second-personal relation – that *constitutes* being owed.

### 5.6 No Accountability for Intentions

The fundamental distinction between the two dimensions of normativity is also expressed in Hobbes's ethics by the fact that each applies to a different range of objects. Rational agents can be responsible, in the attributability sense, not only for actions, but also for mental states: they can have practical reasons to act, but also epistemic reasons to believe and affective reasons to desire. The prudential laws of nature, for example, apply *in foro interno* to all rational agents. Rational agents can be attribution-responsible for at least some of their mental states because they are potentially responsive to reasons via reasoning. By contrast, rational agents can be accountable to others only for voluntary actions – not for purely "*Vitall*" motions, such as breathing, nor for unexpressed mental states. To be accountable to others, two conditions must be met: (1) that for which one is accountable must itself be responsive to reasoning; and (2) others must, at least in principle, be able to hold one accountable for it. Rational agents are accountable only for what is *reason-responsive* but also *accessible* to others.

On the one hand, no one can be accountable – or, indeed, attribution-responsible – for vital motions because they occur involuntarily and hence without any "help of Imagination": they do not inherently express mental states and hence do not answer to reasons. Rational agents can be accountable only for voluntary motions, which "depend alwayes upon a precedent thought," because only voluntary actions express the inner thoughts and desires or intentions that can respond to reasons.<sup>117</sup>

On the other hand, no one can be accountable for unexpressed mental states: "The secret thoughts of a man run over all things, holy, prophane, clean, obscene, grave, and light, without shame, or blame."<sup>118</sup> Because being accountable to others reduces to their having the normative standing to hold one accountable, one cannot *be* accountable for anything unless there is at least one person with standing to hold one accountable. And others cannot have such standing unless it is possible for them in principle to hold one accountable, and hence to discern violations of the relevant expectation. Obligation and accountability are conceptually distinct from motivation and being held accountable, but the former are intrinsically

<sup>117</sup> L 6.1: 78.    <sup>118</sup> L 8.10: 108.

tied to the *possibility* of motivation and being held accountable. Because others cannot access our unexpressed mental states, they could not hold us accountable for them, “for the inward *thought*, and *beleef* of men . . . humane Governours can take no notice of . . . and consequently fall not under obligation.” Again: we can be held accountable only for “actions, and words, for those onely are known, and may be accused; and of that which cannot be accused, there is no Judg at all, but God, that knoweth the heart.” This is why belief in illegal religion is “without blame” when unexpressed: “internall Faith is in its own nature invisible, and consequently exempted from all humane jurisdiction.”<sup>119</sup>

It is true that in arguing that one could not be accountable for unexpressed mental states Hobbes also emphasized that mental states “are not voluntary,” seeming to imply that voluntariness is a further, independent condition of accountability.<sup>120</sup> But as we have seen, unlike for the scholastics, the significance of voluntariness for Hobbes has nothing to do with free will. Its significance is that what is voluntary (1) is preceded by and so expresses one’s thoughts or intentions and is therefore potentially *reason-responsive*, and (2) is outwardly expressed and so is *accessible* to others. Rational agents are potentially accountable only for what is voluntary, but that is because only what is voluntary is both potentially reason-responsive and accessible.<sup>121</sup> Many thoughts and intentions are reason-responsive, but none are accessible to others unless expressed.

It is true that in *Dialogue* Hobbes asserted murder is a felony, whereas homicide is not, because only the former involves “Malice forethought.” As the dialogue’s “philosopher” puts it, “the harm done is the same” in manslaughter (a type of homicide) “as if it had been done by Felony; but the wickedness of the Intention was nothing near so great.”<sup>122</sup> More generally, Hobbes accepted Coke’s definition of a felonious crime as an act committed with *animus felleus*, that is, with malicious intent or “bitterness of mind.” The entire examination of which actions count as “Felony, and Crimes in their own nature without the help of [a particular] Statute” proceeds on the assumption that “the mind of the Offender and the Mischief he intendeth” is what partly distinguishes felonies from non-

<sup>119</sup> L. 40.2: 738; 42.80: 866; 42.43: 822; cf. 27.2: 454.      <sup>120</sup> EL 19.7; DCv 5.8.

<sup>121</sup> Cf. Q: 345/450. Hobbes subscribed to a “rationalist” rather than “voluntarist” criterion of responsibility (Smith 2005), which emerges from the fact he held that one can be attribution-responsible for mental states (see Q: 230/291 discussion of “a just Fear”). EL 25.3 refers only to inaccessibility, not involuntariness.

<sup>122</sup> D: 82.

felonies as well as each type of felony from each other.<sup>123</sup> And it is no use retorting that the *evidence* for one's intentions lies in external actions – evidence is one thing, that of which it is evidence another. What constitutively distinguishes felony from non-felony and each type of felony from another is the intention, not the evidence of the intention.

The appearance that one is thereby accountable for mere intentions is deceptive, however. Hobbes's account, like Grotius's,<sup>124</sup> assumes intentions partly constitute and determine what type of action one has committed: identical behaviours with identical effects but undertaken with different intentions are different types of action. Different types of action, moreover, warrant different punishments. But what follows is that one is accountable to others for *actions*, not merely their *effects*; it does not follow that one is accountable for intentions in themselves. Punishment may be warranted for malicious actions, such as attempted murder, that fail to produce the intended effects, but not for "meer Intentions" however malicious. One is accountable for mental states only insofar as they *constitute* voluntary actions, and accountable for actions only if they are *voluntary* and hence express intentions. Ultimately, one is accountable only for the combination of actions and the attitudes they express. But "of Intentions, which never appear by any outward act, there is no place for humane accusation," and "till it appear by some thing done, or said, by which the intention may be argued by a humane Judge, it hath not the name of Crime."<sup>125</sup> That one is not accountable for mere intentions is precisely why Hobbes ruled out punishment for unexpressed heretical beliefs: it is a grave error "to extend the power of the Law, which is the Rule of Actions onely, to the very Thoughts, and Consciences of men, by Examination, and *Inquisition* of what they Hold, notwithstanding the Conformity of their Speech and Actions."<sup>126</sup>

One category of action may seem to furnish an exception: treason. Treason can "consist in design" or mere intention (to kill the sovereign) "hidden in the Breast,"<sup>127</sup> but Hobbes nevertheless discussed it, in *De Cive*,

<sup>123</sup> D: 78–79. Stoffell (1991) mistakenly thinks the "philosopher" rejects the view that naturally felonious acts possess *animus felleus*. The philosopher rejects Coke's etymology, arguing that some actions without malicious intent have been legislated as felonies *by statute*.

<sup>124</sup> JBP 2.20.18–19.

<sup>125</sup> L 27.2: 452–454. This passage provides the context for interpreting Hobbes's statement, at L 30.13: 530, that subjects "are to be taught, that not onely the unjust facts, but the designs and intentions to do them, (though by accident hindred,) are Injustice; which consisteth in the pravity of the will, as well as in the irregularity of the act." The unjust designs are intentions that find outward expression, without realizing the intended effects. Cf. D: 42.

<sup>126</sup> L 46.39: 1096. <sup>127</sup> L 27.37: 478; D: 73–76. DCv 14.20–22 focusses on outward expression.

*Leviathan*, and *Dialogue*, under the category of “crime.” However, Hobbes did not treat treason as a crime in his technical sense: one is not *accountable* to anyone for treason, which is why when the sovereign tries traitors and puts them to death, he is not inflicting a (legally regulated) *punishment* (in the strict, proper sense) on his *subjects* for a crime under civil law, but is engaged in “acts of Hostility” against “an enemy of the Common-wealth” under “the condition of warre.”<sup>128</sup> This is why Hobbes claimed that outlawing by statute what under natural law already counts as treason is superfluous:<sup>129</sup> it does not add “any [directed] obligation” under “Civill Law.” There exists intrinsically only a “*naturall obligation*” against treason, which is not an obligation in Hobbes’s primary, juridical sense, but simply consists in a reason for which one is accountable to no one.

Nor does the case of God supply an exception to Hobbes’s doctrine. It is true that, although the mere intention to kill is not a crime for which one is accountable to any “humane Judge,” Hobbes wrote that it is still “a sinne,” since “God that seeth the thoughts of man, can lay it to his charge.”<sup>130</sup> He also wrote that even the sovereign, who “is obliged by the Law of Nature” to procure “the safety of the people,” is obliged “to render an account thereof to God, the Author of that law, and to none but him.”<sup>131</sup> But Hobbes held that ‘God’ signifies not our conception of something infinite – we can have no such conception – but of our own incapacity to conceive of anything unlimited. Utterances that deploy the term ‘God’ and seemingly attribute properties to an infinite being merely signify our desire to praise and honour the person of God. We are incapable of using language meaningfully to state that an infinite being has such-and-such natural properties: because “the nature of God is incomprehensible . . . the Attributes we give him, are not to tell one another, *what he is*, nor to signifie our opinion of his Nature, but our desire to honor him with such names as we conceive most honorable amongst our selves,” so that “in the Attributes which we give to God, we are not to consider the signification of Philosophicall Truth; but the signification of Pious Intention, to do him the greatest Honour we are able.”<sup>132</sup> When Hobbes wrote that God “seeth the thoughts of man,” he was expressing himself metaphorically in the language of his contemporaries, not making any descriptive, propositional claims about God’s capacities. He was uttering expressive praise directed to the person of God – what in *Anti-White* he called an oblation, as opposed to a proposition.<sup>133</sup>

<sup>128</sup> L 28.13: 486; 28.23: 494. Cf. DCV 14.22.    <sup>129</sup> L 30.4: 522; DCV 14.20–21.    <sup>130</sup> L 27.2: 452.

<sup>131</sup> L 30.1: 520, italics removed.    <sup>132</sup> L 34.4: 614; 31.33: 568.    <sup>133</sup> AW 35.16.

The person of God could not possibly know other persons' mental states. For according to Hobbes, God is not a natural person. He is an artificial person who does not exist as a person unless represented by someone; like the state, he is an artificial person "by fiction."<sup>134</sup> This is why God has as many *personae* as he has representers,<sup>135</sup> why he is incapable of covenanting without the mediation of a representer,<sup>136</sup> and why he is even incapable of authorizing his own representer.<sup>137</sup> The person of God could know someone's unexpressed mental states only if God's representer could know them, which means God could know the mental states of only his own authorized representer. We therefore have two cases. The first concerns a person who is not God's representer. Such a person cannot be accountable to God for his unexpressed mental states because the person of God could not, via his representer, know them. The second is the case of God's representer. Here God can be said virtually to know, via his representer, the person's mental states. But saying the representer is accountable to God in this virtual sense just amounts to saying he is accountable to himself.

This second case corresponds to that of sovereigns. Because God is not a natural person, his representer must be authorized by someone else, namely, by the very persons for whom he is authorized to represent God. Thus if one asks "Upon what ground" Moses was authorized to represent God to the Israelites, the answer is: by a "submission of their own," i.e., by "the Consent of the People," an act of submission by which the people covenanted to obey the commands uttered by Moses as God's.<sup>138</sup> The one who ultimately represents the person of God to the people of a commonwealth is he who represents the commonwealth itself. The "earthly Sovereign" is rightly called "the Image of God" not because he resembles but because he represents God: he is the ultimate interpreter for his subjects "of what God hath spoken," but also the only one who "can take notice of what is, or what is not the Word of God," i.e., he determines what counts as scripture.<sup>139</sup> Therefore for Hobbes to say the *sovereign* is accountable to God is effectively to say he is accountable to his own conscience. Yet as we shall see in [Chapter 7](#), on Hobbes's account there is no such thing as *intrapersonal* accountability: obligation is always owed to another person. To be "accountable" for one's intentions solely to one's own conscience is to fail to be accountable at all: it is to be only attribution-responsible for them.

<sup>134</sup> Abizadeh (2017a). <sup>135</sup> L 41.9: 772. <sup>136</sup> L 14.23: 210; 18.3: 266. Curley (2004).

<sup>137</sup> Abizadeh (2013: 138–139; 2017a). <sup>138</sup> L 26.40: 446; 40.6: 740. Abizadeh (2013: 142–146).

<sup>139</sup> L 45.17: 1032; 40.4: 738. Pocock (1997) gives a rival account, but see Abizadeh (2013).

*The Laws of Nature, Morality, and Justice*

The distinction between reasons of the good and reasons of the right – and between the attributability and accountability dimensions of normativity – is indispensable to explaining why Hobbes asserted, in chapters 15 and 26 of *Leviathan*, that the laws of nature are not intrinsically laws at all: “These dictates of Reason, men use to call by the name of Lawes; but improperly: for they are but Conclusions, or Theoremes concerning what conduceth to the conservation and defence of themselves; whereas Law, properly is the word of him, that by right hath command over others.” And again: “the Lawes of Nature . . . are not properly Lawes, but qualities that dispose men to peace, and to obedience. When a Common-wealth is once settled, then are they actually Lawes, and not before; as being then the commands of the Common-wealth; and therefore also Civill Lawes.”<sup>1</sup> The so-called laws of nature are not intrinsically laws because, although they articulate the reasons agents have to preserve themselves, no one is intrinsically accountable to others for complying with such reasons: the laws of nature themselves provide “no place for Accusation” – no standing for second-personal, reactive censure.<sup>2</sup> They become true laws, for which persons are accountable, only after embodied in the sovereign’s commands, but until then, violating them is merely to commit “sins of prudence,” not injustice.

It is true that Hobbes followed the initial assertion in chapter 15 by remarking that “yet if we consider the same Theoremes, as delivered in the word of God, that by right commandeth all things; then are they properly called Lawes.”<sup>3</sup> The theological tradition of Hobbes scholarship has seized on remarks such as this as evidence that, for Hobbes, natural laws derive their intrinsic normative force from being, in their essential nature, divine

<sup>1</sup> L 15.41: 242; 26.8: 418. See EL 17.12; DCv 3.33; L 26.16: 430. Hobbes sometimes wrote that “men are commanded” by natural laws (L 14.5: 200), but he also characterized them as “advising [*hortantis*] us” (DCv 3.26). Hobbes called them commands only in the same metaphorical sense in which he called the laws of nature laws.

<sup>2</sup> L 27.3: 454; 14.30: 214; 30.30: 552.    <sup>3</sup> L 15.41: 242. See DCv 3.33.

commands.<sup>4</sup> According to Hobbes, however, not everyone does “consider” the laws of nature “as delivered in the word of God”: atheists do not do so, and theists who deny providence – such as Epicurus or Lucretius – do not do so.<sup>5</sup> Yet the laws of nature are normative for *all* rational agents – for “*every man*”<sup>6</sup> – including atheists and anti-providentialist theists. Whether someone considers the laws of nature to be divine commands is as contingent as whether there exists a civil law incorporating them: neither possibility is intrinsic to what natural law is, or determines to whom it dictates reasons. Acknowledging a providential God in common with others or covenanting to obey a sovereign may give one directed obligations to obey the laws of nature, owed to God’s representer or the sovereign; but such obligations do not arise from any normative force intrinsic to natural laws.

Hobbes’s normative philosophy is therefore split along two dimensions: reasons of the good, represented paradigmatically by prudential natural laws; and reasons of the right, represented paradigmatically by conventional obligations of justice. There is a fundamental chasm between these two dimensions: the latter, accountability dimension is neither reducible to nor derivable from the former, attributability dimension. Whereas natural laws articulate merely prudential reasons, obligations of justice are in an important sense genuinely *moral* obligations.

To this it might be objected that Hobbes on numerous occasions equated natural law to the *moral* law, which includes the third law – namely *justice*. It would therefore seem either that natural laws are not merely prudential but articulate genuinely moral obligations of justice, or that so-called morality and justice in Hobbes amount to nothing more than the agent’s prudential self-interest. The former view dominates interpretations according to which Hobbesian ethics includes a genuinely moral dimension, while the latter view represents the “orthodox” reading according to which Hobbes’s ethics lacks any genuinely moral notion of obligation.<sup>7</sup> In both cases, the normativity of fulfilling conventionally incurred obligations is supposed to be derived from and indeed reducible to the normativity of the third law of nature. Against both currents of scholarship, I shall argue that, whereas natural laws themselves are merely prudential, the conventional obligations of justice are moral in a way that is neither reducible to nor indeed derivable from the normativity of the laws of nature – including the third law.

<sup>4</sup> Warrender (1957: 98 ff.); Martinich (1992). <sup>5</sup> L 31.2: 554–556. <sup>6</sup> L 14.4: 200; 4.8: 52.

<sup>7</sup> For the former, see Lloyd (2009), for the latter, Nagel (1959). Cf. Plamenatz (1965).

None of this is to deny the intimate relation between the two dimensions of normativity in Hobbes. To the contrary, reasons of the good (1) include prudential reasons to *occupy* the second-personal standpoint and to be disposed to *fulfil* the juridical obligations one thereby acquires, and (2) help to determine the *content* of obligations, both negatively (by placing limits on the obligations one can acquire) and positively (by supplementing the intentional content of contractual undertakings). Reasons of the good shape reasons of the right because, on Hobbes's *sign theory of consent*, people acquire obligations only via actions or speech that others reasonably interpret as a sign of their will to acquire such obligations – and reasonably interpreting signs of others' will involves imputing a minimum level of rationality and prudence to their actions.

### 6.1 The Meaning of 'Moral'

The question of whether Hobbes had a genuinely moral notion of obligation is obscured by the fact that what Hobbes meant by 'moral' is different from what moral philosophers typically do today. Contemporary philosophers frequently use the term in one of two ways. First, to denote reasons grounded in a requirement to take others' interests *impartially* into account, with the anti-eudaimonist implication that morality may require acting against one's own interests: prudential and moral reasons in this sense must in principle be able to diverge.<sup>8</sup> Second, to denote reasons for which one is *accountable* to others with standing to demand one's conformity and respond reactively to failures, i.e., reasons of the right.<sup>9</sup> These two conceptions of morality are modern: they arose in the aftermath of the seventeenth century.

Classical writers' use of 'moral' and its cognates was different. In its most generic sense, the Latin term *moralis* was used as the accepted translation for the Greek *ēthikos* to refer to the actions of normatively responsible agents – as Suárez put it, actions meriting "praise or blame."<sup>10</sup> As a result, *mores* could simply refer to the customary norms regulating *social interaction* amongst responsible agents – what in French became 'moeurs', in English 'manners', equivalent to what today we call *social*. Thus to call norms moral in this sense is not even to claim they are genuinely normative, but merely to say they are customarily followed in social interaction.<sup>11</sup>

<sup>8</sup> Nagel (1959: 74); cf. Peters (1956: 174–175); Parfit (1984).

<sup>9</sup> Darwall (2009: 27–29). Cf. Scanlon's (1998: 6–7, 173–176) "narrow" sense of morality.

<sup>10</sup> DLDL VII.1.3. <sup>11</sup> Cf. DLDL VII.1.3–4.

By extension, genuine reasons are moral in this sense insofar as they govern the passions or actions involved in social interaction. Here, moral reasons neither require impartially considering others' interests, nor are owed to others; the mere fact they govern the passions or actions involved in social interaction, and foreseeably *affect* others' interests, is sufficient to call them moral.

Beyond this social-interactionist sense, classical authors also used 'moral' more strictly to mean what serves the *common good*. This is the sense in which classical theorists took natural laws to be moral laws: they are oriented to the common good – indeed, intrinsically oriented to it. This is not to say moral reasons in this sense are *grounded* in a common good (the way the impartialist sense grounds reasons in others' interests): as we saw in the introduction, for the eudaimonistic classical theorists, practical reasons are all grounded in one's own good; the common good is a constituent of each person's own good.

Hobbes used the term 'moral' in the two senses he inherited from classical writers. He rather clearly announced the social-interactionist sense the first time he equated natural laws to moral laws in *Elements*: "they are called the Lawes of Nature, for that they are the dictates of naturall reason," but are called "also Morall Lawes, because they concerne mens manners, and Conversation one towards another."<sup>12</sup> ('Conversation' in its seventeenth-century sense means social interaction.) And in *De Cive*, he linked the social-interactionist to the common-good sense: the reason why "Legem naturalem" may also be called "*lege morali*," Hobbes argued, is that they regulate the potential outbreak of *discord in social interaction*: they prescribe "the necessary means to Peace," which consist in "*Good Manners* [*bonos mores*]," "*habits*," or "*Vertues*."<sup>13</sup> Yet Hobbes's nod to the common-good sense of 'moral' departs from his classical predecessors: natural laws are moral not because they are *intrinsically* oriented to the common good of peace – they are intrinsically oriented to self-preservation – but because peace happens to be *instrumental* to self-preservation.

In *De Cive*, Hobbes produced a more complicated classification of natural laws that reinforces the social-interactionist meaning of 'moral'. The first complication lies in a new contrast he drew between laws of nature concerning the relevant *social* means of self-preservation and those concerning its relevant *non-social* means. Insofar as the doctrine of *De Cive* is concerned with "*Philosophiæ moralis*," and insofar as the "*leges naturales*" amount to moral philosophy, Hobbes declared that "I have onely delivered

<sup>12</sup> EL 18.1.    <sup>13</sup> DCv 3.31.

such precepts in this place, as appertain to the preservation of our selves against those dangers which arise from discord.”<sup>14</sup> The laws of nature amounting to moral laws do not concern just *any* type of means for self-preservation; they concern the relevant *social* means of self-preservation, which is why the foundational moral law prescribes peace in particular. However, having defined natural law more broadly as “the Dictate of right Reason, conversant about those things which are either to be done, or omitted for the constant preservation of Life, and Members,”<sup>15</sup> Hobbes did recognize that reason also prescribes the relevant *non-social* means of self-preservation: “there are other precepts of *rationall* nature, from whence spring other vertues: for *temperance* also is a precept of Reason, because intemperance tends to sicknesse, and death. And so fortitude too . . . because it is a means tending to the preservation of him that resists.”<sup>16</sup> Virtues concerning the non-social means of self-preservation do not feature in the doctrine of *De Cive* because, although they are natural laws broadly understood, they are not *moral* laws. For the “legi morali,” Hobbes insisted, “consists . . . in the prohibition of . . . offences, whereby our Neighbours are *laeduntur* [hurt, wounded, offended].”<sup>17</sup> Indeed, in the 1647 second edition, Hobbes clarified that the definition of natural law, in the (moral) sense at stake in *De Cive*, appeals to a narrow sense of right reason restricted to reasons governing social interaction: “By *Rectam rationem* . . . I understand . . . the peculiar and true ratiocination of every man concerning those actions of his which may either redound to the dammage [*damnum*], or advantage [*utilitatem*] of his neighbours.”<sup>18</sup>

Hobbes made the same point when he introduced a second point of contrast, between the “secular” laws of nature, which concern the relevant means of self-preservation, and the “sacred” laws of nature, which are dictates of reason governing the proper way to “honour and worship the Divine Majesty.” Here he again clarified that the former “dictates of reason,” or the “*naturall Lawes*” comprising specifically “*Morall vertues*,” are those that “pertain to the discharge of the duties of men one toward the other [*quæ pertinent ad officia hominum erga se inuicem*].”<sup>19</sup> These are duties “toward” others not in the sense of being *owed* to them, but again in the sense that they regulate social interactions *affecting* each other.<sup>20</sup> As Hobbes reiterated in the same chapter, whereas the sacred laws of nature comprise “the naturall Lawes concerning the worship of God,”

<sup>14</sup> DCv 3.22.    <sup>15</sup> DCv 2.1.    <sup>16</sup> DCv 3.22.    <sup>17</sup> DCv 18.3.    <sup>18</sup> DCv 2.1.    <sup>19</sup> DCv 15.8.

<sup>20</sup> The same holds for the term ‘erga’ in a note added to *De Cive*’s 1647 edition, which states that the (moral) laws of nature consist in the “*officia sua erga cæteros homines ad conservationem propriam necessaria*.” DCv 2.1.

the secular moral laws are “those which concern justice, and the carriage of men towards men [*mores hominum versus homines*].”<sup>21</sup>

Hobbes zeroed in on the common-good meaning of ‘moral’ in *Leviathan* with ever greater precision, specifying that moral philosophy concerns the kind of social interaction implicating *social discord*: the interest at stake in moral philosophy is the common good of peace. Drawing a distinction between etiquette or “*Small Moralls*” and the type of morals at stake in moral philosophy, and again deploying the English term ‘manners’ for the Latin *mores*, he wrote: “By MANNERS, I mean not here, Decency of behaviour; as how one man should salute another, or how a man should wash his mouth, or pick his teeth before company, and such other points of the *Small Moralls*; But those qualities of man-kind, that concern their living together in Peace, and Unity.”<sup>22</sup> Thus “Morall Philosophy” in Hobbes’s technical sense of the phrase “is nothing else but the Science of what is *Good*, and *Evill*, in the conversation [i.e., interaction], and Society of man-kind.” This is why “the true Doctrine of the Lawes of Nature” or “*Morall Vertues*” is “the true Morall Philosophie”<sup>23</sup> – indeed, the entirety of moral philosophy in Hobbes’s technical usage. This is also the sense in which natural laws “are called not onely *Naturall*, but also *Morall* Lawes; consisting in the *Morall Vertues*,” that is, “all habits of the mind that conduce to Peace, and Charity.”<sup>24</sup> Repeating again his distinction between the social and non-social means of self-preservation, Hobbes reaffirmed that only the former are at stake in moral philosophy. Moral philosophy in Hobbes’s sense comprises only those

Lawes of Nature, dictating Peace, for a means of the conservation of men in multitudes; and which onely concern the doctrine of Civill Society. There be other things tending to the destruction of particular men; as Drunkenness, and all other parts of Intemperance; which may therefore also be reckoned amongst those things which the Law of Nature hath forbidden; but are not necessary to be mentioned, nor are pertinent enough to this place.<sup>25</sup>

Finally, reiterating *De Cive*’s distinction between the sacred and the secular moral dictates of reason, Hobbes characterized the former as the “Dictates of Natural Reason . . . touching the Honour and Worship of the Divine Majesty,” and the latter as concerning “the naturall Duties of one man to

<sup>21</sup> DCv 15.15; 15.17.    <sup>22</sup> L II.I: 150; LL 15: 151.    <sup>23</sup> L 15.40: 242.    <sup>24</sup> L 26.36: 442.

<sup>25</sup> L 15.34: 238. Cf. LL 15: 239; L 31.40: 572.

another” – again, not in the sense of being owed to others, but in the sense that they regulate social interaction.<sup>26</sup>

If the (social) laws of nature exhaust the content of moral philosophy, then moral philosophy in Hobbes’s sense concerns only prudential reasons governing the dispositions and actions that are the relevant means for peaceful social interaction. Moral philosophy in Hobbes’s sense therefore does not concern moral reasons in either of the modern senses outlined above. Because natural laws are self-regarding, moral philosophy in Hobbes’s sense does not encompass reasons that are moral in the impartialist sense; and because moral philosophy does not encompass juridical obligations arising by convention, it does not concern moral reasons in the accountability sense either.

This does not mean, however, that Hobbes’s normative philosophy fails to include a genuinely moral philosophy in one of our contemporary senses. It fails to include genuinely moral obligations in the impartialist sense, but does include them in the accountability sense. To be sure, Hobbes himself did not use the term ‘moral’ to denote the accountability dimension of normativity; nor did he consider this dimension to be an intrinsic element of what he meant by ‘moral philosophy’; but the dimension is nevertheless present in, indeed is fundamental to, Hobbes’s normative philosophy. His own terms of art for it were, not ‘morality’, but ‘injury’, ‘wrong’, and ‘justice’.

## 6.2 Reasons of the Right Cannot Be Derived from Reasons of the Good

I have contended that Hobbesian natural laws are moral in the traditional *social-interactionist* and *common-good* senses but not in the modern *impartialist* and *accountability* senses. This contention might be challenged by noting that Hobbes characterized the laws of nature as precepts of *reciprocity*, which can all be

contracted into one easie sum . . . *Do not that to another, which thou wouldst not have done to thy selfe*; which sheweth him, that he has no more to do in learning the Lawes of Nature, but, when weighing the actions of other men with his own, they seem too heavy, to put them into the other part of the ballance, and his own into their place, that his own passions, and self-love, may adde nothing to the weight; and then there is none of these Lawes of Nature that will not appear unto him very reasonable.<sup>27</sup>

<sup>26</sup> L 31.7: 560. *Leviathan* does not explicitly call the sacred dictates of reason “natural laws.”

<sup>27</sup> L 15.35: 240; see 14.5: 200; 26.13: 424; 27.4: 456; 42.11: 786.

This has suggested to some that natural laws are, rather than merely prudential precepts of self-preservation, genuinely moral laws in the modern, impartialist sense. S. A. Lloyd has recently argued, for example, that Hobbes's laws of nature specify natural duties requiring "impartial attention to the interests of humanity." They are moral laws in the impartialist sense because their foundation is a "fairness" principle of "reciprocity." Their normative ground lies not in the individual's self-preservation, but in "the good of *humanity*."<sup>28</sup>

Lloyd recognizes that Hobbes characterized the laws of nature as dictates of reason prescribing the relevant means to self-preservation, but she points to a passage from *De Cive* as counterevidence in favour of her impartialist interpretation: reason dictates that each "must foregoe that Right [of nature] for the preservation of mankind [*conseruationem humani generis*]."<sup>29</sup> This is a rather anaemic basis for the impartialist interpretation, however, because whereas the many passages grounding natural law in individual self-preservation are incompatible with the impartialist reading, this *De Cive* passage is compatible with the self-preservationist reading. The fundamental law of nature is to seek peace precisely *because* seeking peace – preserving humankind – is the relevant social means for each individual's *self*-preservation.

The weight of Lloyd's interpretation is therefore carried by her observation that, while acting to promote peace is *in general* the relevant means to self-preservation, under some circumstances it is not: the actions normally conducing to peace may under some circumstances undermine self-preservation. Lloyd draws two conclusions from this possible divergence: first, that under such circumstances, the laws of nature, in prescribing peace and its relevant means, thereby prescribe actions contrary to the agent's means of self-preservation; second, that the impartialist reading can account for this putative feature of natural law, but the self-preservationist reading cannot. The upshot is supposed to be that the latter reading must deny such circumstances could ever arise, holding instead that the "Laws of Nature must, and so do, *always* secure the interest of the agent who is to follow them."<sup>30</sup>

But the self-preservationist reading is not committed to denying the obvious fact – which Hobbes explicitly recognized in his treatment of the Foole, capital punishment, and, of course, the state of nature – that

<sup>28</sup> Lloyd (2009: 138, 14, 110, 98). Lloyd (2009: 102–103) thinks this means that natural law is moral in the accountability sense as well, and imposes duties that "men owe to one another." But this results from Lloyd's conflation of the distinct senses of 'morality' distinguished earlier.

<sup>29</sup> DCv 15.5, quoted in Lloyd (2009: 112). <sup>30</sup> Lloyd (2009: 114).

sometimes the actions necessary to preserve oneself will not be conducive to peace. What is Lloyd's warrant for claiming the self-preservationist reading denies this? The alleged warrant lies in her first conclusion, namely that under some circumstances the laws of nature prescribe actions noxious to one's own preservation. But that first conclusion is false. Lloyd draws it only because she misinterprets Hobbes's claim that natural laws always oblige *in foro interno* but not always *in foro externo*. Lloyd's misinterpretation shows up in her formulation of an objection she considers, on behalf of the self-preservationist reading, against her own. The objection is that natural laws are not always normative for rational agents (which is, of course, the Foole's view). The objection's premise, on Lloyd's telling, is that according to Hobbes "an individual is not required to do the actions dictated by the Laws of Nature whenever it would 'procure his own certain ruin' to do so." Lloyd retorts that although there "is no question" the premise is true, it "does not follow" from it that natural law directs "individuals to preserve themselves," or that it is "justified by reference to the interests of the agent." The premise merely specifies a fact "about the conditions under which it is suspended," not about its content or grounds.<sup>31</sup>

In fact the premise, which Lloyd endorses, is false, and the self-preservationist reading is not premised on it. Lloyd accordingly misstates the relevant objection. The relevant objection is that natural laws *do not ever* "dictate" or prescribe actions that would procure one's own "certain," and hence reasonably knowable, ruin. Lloyd is mistaken to claim that sometimes the dictates of natural law are "suspended."<sup>32</sup> Hobbes's point, rather, is that for someone to undertake actions that would "make himselfe a prey to others, and procure his own certain ruine, [is] contrary to the ground of all Lawes of Nature, which tend to Natures preservation."<sup>33</sup> Contrary to Lloyd's assertion, Hobbes identified the "ground" of natural law with each agent's own "preservation." When Hobbes asserted that, while the laws of nature always oblige *in foro interno*, they do not always oblige *in foro externo*, his point was that they always prescribe the *dispositions* that are the relevant means to promoting peace, but do not always prescribe the *actions* that *would be* the means for promoting peace if it were safe for one to undertake them. It is not a law of nature to *act* justly, gratefully, etc., when it is reasonably knowable or foreseeable that doing so would lead to self-destruction.

<sup>31</sup> Lloyd (2009: 115–116).

<sup>32</sup> For the same mistake, see Murphy (1995).

<sup>33</sup> L 15:36: 240.

That is precisely what the self-preservationist reading, but not the impartialist reading, correctly asserts.<sup>34</sup>

Reciprocity, which Hobbes cited as the “core” or “sum” of the law of nature, is not an impartialist ground from which natural laws are derived. Reciprocity has precisely the same *prudential* normative status as the virtues articulated in Hobbes’s list of natural laws: being disposed to adhere to reciprocity is a relevant social means for self-preservation, because it is a relevant social means for peace. That is why it serves as a convenient heuristic device.

A critic might concede that natural laws are not moral in the impartialist sense but nevertheless insist they cannot be reduced to self-regarding reasons of the good because they are moral in the *accountability* sense. Hobbes, after all, included *justice* – which, I have argued, is his term par excellence for the second, accountability dimension of normativity – as the third law of nature. The inclusion of this law, “*That men performe their Covenants made,*”<sup>35</sup> might suggest the normativity of covenantal obligations is grounded in or amounts to nothing more than the normativity of the third law of nature,<sup>36</sup> and that if, as I have argued, covenantal obligations are directed obligations, then the third law of nature must also articulate a directed obligation for which one is accountable. This challenge is illusory. Just as it is prudent to be disposed to reciprocity in general, it is also prudent on Hobbes’s view to be disposed to live up to one’s directed obligations in particular. This is why, given that “every subject in a Common-wealth, hath covenanted to obey the Civill Law,” the positive law for which subjects are accountable becomes, in virtue of the third law of nature, “part also of the Law of Nature” prescribing their own good.<sup>37</sup> When one makes a covenant, it is *both* the case that one has a reason of the good to be disposed to keep it, as dictated by the third law of nature, *and* the case that one has a reason of the right to keep it, for which one is accountable to the obliger.<sup>38</sup> The point is not that the two dimensions of normativity cannot

<sup>34</sup> Lloyd (2009: 54, 147) also asserts that the natural law added in *Leviathan’s* conclusion – “*That every man is bound by Nature, as much as in him lieth, to protect in Warre, the Authority, by which he is himself protected in time of Peace*” (L R&C.5: 1133) – counts as further evidence against the self-preservationist reading, because it is not “remotely plausible” that “every individual in every case stands to profit by participating in war to defend his besieged sovereign.” But Hobbes qualified this law in the same way he qualified other precepts: the qualification “as much as in him lieth” indicates that it does not prescribe what could reasonably be known or foreseen to lead to one’s ruin.

<sup>35</sup> L 15.1: 220. <sup>36</sup> Murphy (1995: 854). <sup>37</sup> L 26.8: 418.

<sup>38</sup> Suárez’s claim that God’s command superimposes juridical obligation on the intrinsic normativity of what is necessary to felicity is analogous. DLDL II.6.11–12.

overlap – they can, and here they do – but that the one is not reducible to or derivable from the other.

To all of this it might be objected that although reasons of the right are not, for Hobbes, *reducible* to reasons of the good, they are nevertheless constitutively *grounded* in or *derivable* from them. We can flesh out this objection by focussing on Hobbes's status as a virtue theorist. Hobbesian laws of nature universally prescribe *in foro interno* a *disposition* to act in ways conducive to peace, but only circumstantially prescribe *in foro externo* the *actions* conducive to peace.<sup>39</sup> David Boonin has on this basis argued that Hobbes was a virtue theorist and, as such, had a genuine moral theory that does not collapse into mere prudentialism. True, the *justification* for cultivating virtue or being virtuous is prudential: it consists in the fact that being disposed to act virtuously better promotes one's felicity. The *content* of the virtues, however, refers not to the agent's own good, but to the social good of peace: virtuous agents do not act "directly out of a concern for prudence" or their own good, as Boonin puts it, but directly out of a concern to act virtuously (peacefully) in their interactions. Hobbes's theory is, he argues, *indirectly* but not directly prudential.<sup>40</sup> This is correct: as I argued in Chapter 4, Hobbes did not take one's own felicity to be the proper *aim* of each action.

One might now argue as follows. On the one hand, the second law of nature prescribes that individuals (be disposed to) covenant with others for the sake of peace; therefore, the *content* of natural law effectively prescribes that individuals occupy the second-personal standpoint by taking on directed obligations. Virtuous persons who take on obligations will therefore be disposed to act justly out of a concern to uphold their obligations. On the other hand, this prescription is for each individual ultimately *justified* by, and its normative force *derived* from, the premise that being disposed to occupy the second-personal standpoint serves the individual's own interests.<sup>41</sup> One might therefore conclude, contrary to my thesis, that for Hobbes, although reasons of the right are not *reducible* to reasons of the good (their content is different), the former are *derivable* from the latter (they are indirectly justified prudentially).

Yet this argument does not show reasons of the right to be grounded in reasons of the good. There is a difference between the normative reason rational agents have to occupy the second-personal standpoint and the normative justification for that standpoint itself. According to Hobbes, we

<sup>39</sup> L 15.36: 240; 15.39: 240; 26.8: 418; 26.36: 442; DCv 3.29–31.      <sup>40</sup> Boonin-Vail (1994: 62–64).

<sup>41</sup> See LeBar's (2009: 662–669) criticisms of Darwall (2009: 59). Cf. Railton (1997).

have prudential reasons to (be disposed to) take on directed obligations (second law) and to fulfil them (third law). But nothing in the law of nature itself shows we have the normative *power* to take on such obligations. The argument conflates (a) the prudential justification offered to rational agents to occupy the second-personal standpoint with (b) the justification for claiming they have the normative power to take on directed obligations and hence to occupy that standpoint. Showing that it would be good for agents to acquire or to have directed obligations does not show they have (the normative power to acquire) such obligations. (Showing it would be good for us if a benevolent God existed does not show he exists or that we can create him.)

For Hobbes, reasons of the right are implicitly grounded not in the foundationally normative Prudential Precepts, but in foundational normative precepts distinctive of the second dimension of normativity. Agents do not have the power to occupy the second-personal standpoint and take on obligations *qua* rational agent: they have such a power only if they are *persons* recognized by other persons. The Juridical Precepts for action (as we might call them) prescribe that one  $\phi$  if one has acted such that another person P can reasonably take, and does take, one's previous actions to be a sign of one's will to communicate the intention to acquire a reason to  $\phi$ , and P accepts this will as binding. They also prescribe that, if P subsequently judges that one has violated the putative reason and so responds reactively, one offer a normative response accounting for one's action (and that no one else respond reactively to P's reactive response). If Hobbes's commitment to the normativity of the Prudential Precepts constitutes his *prudentialism*, then his view that something like the Juridical Precepts are the only foundationally normative precepts of the right constitutes his *voluntarism* and *interpersonalism* with respect to juridical obligations. The reasons of the right articulated by Hobbes's account of juridical obligations therefore neither reduce to nor are derived from reasons of the good. The challenge is to see what relation does hold between these two dimensions of normativity in Hobbes.

### 6.3 The Relation between Natural Law and Obligation

Although the prudential imperatives of self-preservation cannot ground directed obligations, they do, on Hobbes's account, (1) yield prudential reasons (a) to *occupy* the second-personal standpoint from within which obligations are acquired and (b) to be disposed to *fulfil* those obligations; they also (2) help *determine* the content of the obligations that can be

acquired, both (a) negatively *limiting* the range of eligible contents, and (b) positively *supplementing* the propositional content of obligation-producing conventions. These two relations are crucial to Hobbes's ethics. For the fact that the two dimensions of normativity have distinct grounds broaches the possibility that juridical obligations could, at least in principle, require actions contrary to one's own good. But Hobbes had to immunize his ethics from this possibility because, for Hobbes, agents' reasons must be not only epistemically but also motivationally accessible to them: nothing could be a normative reason for us that could not in principle be a motivational reason. Since on Hobbes's egoist-hedonist psychology all action stems from desire, and to desire an object is to see it as good for oneself, if juridical obligations required actions we could not see as good, they could not be normative for us. Thus although our good is not the normative *ground* of our juridical obligations, they must nevertheless coincide with our own good.

The concept that enabled Hobbes to articulate the two relations between reasons of the good and right is the pure state of nature. The *pure* state of nature is an ahistorical construct abstracting away from relations of accountability and justice: as Hobbes put it in *Elements*, when "Considering men" in such a state, we consider them "without Covenants, or Subjection one to another, as if they were but even now all at once created male, and female."<sup>42</sup> He reiterated the fantastical character of this pure state in *De Cive*, inviting his readers to "return again to the state of nature, and consider men as if but even now sprung out of the earth, and suddainly (*like* Mushromes) come to full maturity without any kind of obligation to each other [*sine omni unius ad alterum obligatione*]."<sup>43</sup> When considering humans "in the pure state of nature [*in statu merè naturali*]," we imagine them to be in a state "before such time as men had engag'd themselves by any Covenants, or Bonds [*ullis pactis sese invicem obstrinxissent*]."<sup>44</sup> In all three political treatises, Hobbes described this pure state as one in which by nature "every man has a Right to every thing"<sup>45</sup> and hence is devoid of directed obligations. In a 1647 addition to *De Cive*, Hobbes clarified the heuristic nature of the "pure state" by asserting that, in reality, "a Sonne cannot be understood to be at any time in the State of Nature" in relation to a parent who protects him.<sup>46</sup> This is why the thought experiment devoid of obligations requires

<sup>42</sup> EL 22.2.    <sup>43</sup> DCv 8.1.    <sup>44</sup> DCv 1.10.

<sup>45</sup> L 14.4: 198. For this phrase, see also L 15.2: 220; 28.2: 482. Cf. EL 14.10; 15.3; 17.2; 20.2; 27.8; DCv 1.10.

<sup>46</sup> DCv 1.10.

(counterfactually) imagining humans as having immediately sprung “out of the earth” as fully formed adults. The actual, historical state of nature is never pure in this way. It is possible – indeed, for survival it is necessary – to acquire conventional obligations in the state of nature. In the actual, historical state of nature, what is missing is not directed obligations, but the specifically *political* obligation to obey a sovereign. The absence of political obligation – not directed obligations in general – is the defining feature of the historical state of nature.<sup>47</sup>

If the role of the historical state of nature in Hobbes’s theory is to isolate the causes and consequences of war, the purpose of the counterfactually imagined, pure state of nature is to articulate the rational motivation, limits, positive content, and normative grounds of conventional obligations. The thought experiment enables him, for example, to articulate the second law of nature “*to lay down this right to all things.*”<sup>48</sup> To be sure, part of the law’s point is to articulate the prudential reasons for occupying the *position of a civil subject*: one exits the historical state of nature by transferring the relevant parts of one’s right of nature to a sovereign, and the second law asserts that doing so is a relevant means of self-preservation. But one also lays down parts of one’s unrestricted natural right anytime one takes on any conventional obligation, even in the state of nature. A vital example is the state-of-nature obligations owed to fellow “Confederates.”<sup>49</sup> Hence another point of the second law of nature is to articulate the prudential reasons for occupying the *second-personal standpoint* as such. Humans not only have prudential reasons to be disposed “to keep” their covenants, but also “to make” covenants in the first place.<sup>50</sup>

Prudential reasons also place *limits* on the second-personal standpoint, constraining the valid content of directed obligations. Hobbes’s prudentialism commits him to the view that agents have reasons of the good to take the relevant means to ongoing self-preservation – in the broad sense of preserving a life worth living. Hobbes also held that one can never become obligated to do what could be reasonably expected to precipitate self-destruction – in the broad sense of destroying a life worth living. If fulfilling the terms of any conventional act of renunciation, donation, or contract could be reasonably known or foreseen as likely to hasten death or inflict ongoing, abject misery, then the convention is void and the

<sup>47</sup> Lloyd (1992). See Hobbes’s discussion of the historical state of nature in the Americas, interstate relations, and civil war. L 13.11–12: 194–196.

<sup>48</sup> L 14.5: 200.    <sup>49</sup> L 15.5: 224.    <sup>50</sup> L 15.4: 222.

corresponding obligation null. These prudential limits on obligation operate in two ways: globally and locally.

For some putative obligations, it could be reasonably known or foreseen that, regardless of circumstances, fulfilling them would likely precipitate self-destruction. Hobbes ruled out terms like these, and characterized these global limits as the inalienable components of the right of nature. These limits can be seen most clearly in the covenant to obey the sovereign. Despite subjects' covenant to defer, in their actions, to the sovereign's public judgement and to obey his commands, each retains the liberty-right to disobey the sovereign when it can be reasonably known or foreseen that obedience would likely be noxious to self-preservation. And because 'self-preservation' is here understood in the broad sense encompassing not just the bare preservation of a life but of a life worth living, there are fundamentally two types of inalienable liberty-right retained from the right of nature: (1) the *right to bare survival* and (2) the *right to a life worth living*.<sup>51</sup>

The inalienable right to survival consists in a subject's right to disobey his sovereign's command if it is something he could reasonably know or foresee would precipitate his death. It paradigmatically includes the right to disobey any command "to kill, wound, or mayme himselfe; or not to resist those that assault him; or to abstain from the use of food, ayre, medicine, or any other thing, without which he cannot live."<sup>52</sup> To these paradigmatic instances of the right Hobbes added secondary ones. After asserting that each subject retains "the right of resisting them, that assault him by force, to take away his life," he added that the "same may be sayd of Wounds, and Chayns, and Imprisonment."<sup>53</sup> He also cited a right to refuse "to accuse ones selfe, without assurance of pardon": subjects retain this right against self-incrimination because "the Accusation is followed with Punishment; which being Force, a man is not obliged not to resist."<sup>54</sup> This fuller list can be understood to instantiate the right to survival because wounds, chains, and imprisonment – not to mention capital punishment – might be reasonably expected (in the seventeenth century) to precipitate death; the right to resist assaults aimed at wounding, chaining, or imprisoning is justified in part "because a man cannot tell, when he seeth men proceed against him by violence, whether they intend his death or not."<sup>55</sup>

Yet Hobbes also immediately followed this justification with another: the motive for covenanting to obey a sovereign is not just "the security of

<sup>51</sup> For a helpful classification of the liberty-rights retained in the commonwealth, see Sreedhar (2010: chapter 2).

<sup>52</sup> L 21.12: 336. Cf. L 14.29: 214; DCv 2.18.   <sup>53</sup> L 14.8: 202.   <sup>54</sup> L 14.30: 214; cf. 21.13: 338.

<sup>55</sup> L 14.8: 202.

a mans person, in his life,” but also the security of “preserving life, as not to be weary of it.”<sup>56</sup> In other words, subjects retain a right to defend themselves against wounds, chains, and imprisonment also because being free from such harms, even while alive, is amongst the essentials of a life worth living.<sup>57</sup> These latter rights therefore also instantiate the second type of right each subject retains: the inalienable right to a life worth living, and so the right to disobey a command to do anything the subject could reasonably know or foresee would lead to an odious life. After asserting that each enjoys a right not to “accuse ones selfe, without assurance of pardon,” Hobbes then claimed “The same is also true, of the Accusation of those, by whose Condemnation a man falls into misery; as of a Father, Wife, or Benefactor.”<sup>58</sup> A life of “misery” is, of course, a life not worth living. Since survival is not always at stake in the cases covered by these rights, many readers have puzzled over Hobbes’s inclusion of them. But Hobbes was perfectly consistent on this score; his readers have been misled by the false belief that the ultimate ground of the right and laws of nature is bare survival. The point becomes crystal clear when we attend to the even starker case Hobbes contemplated in *De Cive*: a subject has the right to disobey his sovereign’s command “to execute a Parent,” since “a Son will rather die, then live infamous, and hated [or: in infamy and loathing (*infamis atque exosus*)]” for having done so.<sup>59</sup> To be a son who obeyed such a command is to have an intense, lifelong, and continuously frustrated aversion to being the person one has become: one’s life would become so loathsome as to be worth ending.<sup>60</sup>

The question is how reasons of the good constrain reasons of the right in this way. The answer has a systematic basis in Hobbes’s philosophy. His argument for inalienable rights is grounded in his theory of speaker-meaning and *sign theory of consent*.<sup>61</sup> The theory’s key feature is that Hobbes did not, like contemporary liberals,<sup>62</sup> take actually consenting to consist in actions undertaken with the *actual* intention to (communicate the intention to) take on obligations. For Hobbes, consenting or contracting to take on obligations consists instead in actions undertaken by a person (or his representer) that others reasonably take to be a *sign* that

<sup>56</sup> L 14.8: 202. <sup>57</sup> Sreedhar (2010: 68). <sup>58</sup> L 14.30: 214. Cf. DCv 2.19.

<sup>59</sup> DCv 6.13. Hobbes added that “There are many other cases, in which, since the Commands are shamefull [*inhonest*] to be done by some, and not by others, Obedience may, by Right, be perform’d by these, and refus’d by those.”

<sup>60</sup> “life it selfe with the condition of Indureing Scorne, is not esteemed worth the Injoying.” EL 16.11.

<sup>61</sup> On expression-meaning and speaker-meaning, see Grice (1968). Cf. Abizadeh (2015).

<sup>62</sup> Simmons (1979).

he wills to take on the obligations. Some rights are inalienable because, even if some persons *actually* willed or intended to give these rights up, they could never produce signs that others could reasonably take to signify such a thing. Others could never reasonably *impute* such a will or intention to them: there are “some Rights,” namely the rights to survival and a life worth living, “which no man can be understood by any words, or other signes, to have abandoned, or transferred.”<sup>63</sup> Hobbes was not making a point about the conceptual or psychological impossibility of willing or intending self-destruction or self-destructive actions. He was perfectly aware that such intentions or actions are both conceptually and psychologically possible. Sometimes – albeit rarely – one may even have a sufficient reason for self-destructive action. Hobbes was, rather, making a point about speaker-meaning.<sup>64</sup> The intersubjective meaning of people’s utterances and actions, according to Hobbes, is in part determined by applying canons of prudence and rationality: the meaning of conventional signs is constituted not by speakers’ actual will, but by the will that can be reasonably imputed on the presumption that they act in a minimally prudent and rational fashion.

Hobbes also held that some terms of obligation are ruled out by prudential considerations under particular circumstances. Local limits on obligation may arise, for example, in state-of-nature covenants. After a person undertakes an initially valid covenant in the state of nature, any suspicion of his, arising from some “new fact, or other signe,” that the other parties intend to defect after he performs his part, is automatically a “reasonable suspicion,” and is therefore reasonable grounds to “feare” that to perform first is to “betray himselfe to his enemy.” This renders the covenant “Voyd.” (This is why valid covenants in the state of nature, while possible, are so precarious. But “if there be a common Power set over them both,” then his “feare is no more reasonable.”)<sup>65</sup>

By contrast, someone who, on his own initiative, “inrowleth himselfe a Souldier, or taketh imprest mony, . . . is obliged” in virtue of his soldier contract “not only to go to the battell, but also not to run from it, without his Captaines leave.” The obligation to obey his captain is not voided by the soldier’s fear, because his voluntary enrollment or acceptance of

<sup>63</sup> L 14.8: 202.

<sup>64</sup> This points to a flaw in Sreedhar’s interpretation. Sreedhar correctly argues it is neither conceptually nor psychologically impossible to will self-destructive actions, but fallaciously concludes that Hobbes (contrary to L 14.8) took the right to survival to be alienable and to have been alienated in the “soldier contract” (2010: 28–40, 84–86).

<sup>65</sup> L 14.18: 210.

payment “taketh away the excuse of a timorous nature.”<sup>66</sup> The soldier’s voluntary enrollment signifies to *others* that any eventual fears will not undermine his will to soldier on. As this soldier example indicates, Hobbes did not hold that one never has prudential reasons to take *risks* or to take on risky obligations via contract: one may even have a decisive reason to risk one’s life, “to execute any dangerous . . . Office.”<sup>67</sup> Sometimes one increases the odds of self-preservation by taking a lower short-term risk to avoid a greater long-term one.<sup>68</sup> If the long-run odds of self-preservation are increased by risking one’s life in battle to defend one’s commonwealth, then one may have decisive prudential reasons to do so. Hobbes’s view was that one never has decisive prudential reasons to take *bad* risks: to undertake actions or take on obligations that one could reasonably know or foresee would likely hasten death or inflict ongoing misery. The soldier example also reiterates the fact that Hobbes’s notion of a *reasonable* expectation that some action will precipitate death is not reducible to the *actual* fear it will do so: prudential reasons for Hobbes depend neither on subjective beliefs about outcomes, nor on actual outcomes, but on the probable outcomes that are reasonably knowable or foreseeable *ex ante*.

The example does not support, however, the thought that a voluntarily enrolled soldier’s prudential reasons place no limits on his directed obligations. Even a voluntary soldier – contrary to recent commentary – “can never abandon” the right of “defending his life.”<sup>69</sup> His obligation to obey is void as soon as he can reasonably foresee that obeying would precipitate his death:

where a number of men are manifestly too weak to defend themselves united, every one may use his own reason in time of danger, to save his own life, either by flight, or by submission to the enemy, as hee shall think best; in the same manner as a very small company of souldiers, surpris’d by an army, may cast down their armes, and demand quarter, or run away, rather than be put to the sword.<sup>70</sup>

True, a “company of souldiers” is by definition led by a captain; but the passage makes it clear that in such circumstances each individual “may use his own reason” and does not owe it to his sovereign to await his captain’s permission. It is therefore not merely a discernible “new fact” about the

<sup>66</sup> L 21.15: 338–340.    <sup>67</sup> L 21: 15: 338.    <sup>68</sup> Hampton (1986: 16).

<sup>69</sup> L 14.18: 210. See Baumgold (1983); Sreedhar (2010). If there is an exception to the inalienability of the right to survival, it lies not with soldiers but with pastors charged with preaching to infidels (because being willing to die for *non-earthly* felicity might appear potentially compatible with imputed reasonability). AB: 321; Olsthoorn (2014: 167).

<sup>70</sup> L 20.15: 314. See L R&C.6: 1134.

suspicious intentions of other parties that can void a covenant, but also new facts concerning the reasonably knowable or foreseeable consequences of fulfilling the covenant in general.

Because obligations arise from interpersonal signs of the will, prudential reasons also help *positively* to determine their content. With respect to “the act of our *Submission*” in sovereignty-covenants, for example, one part of the content of “our *Obligation*” is “drawn from the expresse words” themselves, but another part is inferred “from the Intention of him that submitteth himselfe to his Power.” The “Intention” in question is, once again, not the person’s *actual* intention, but “is to be understood by the End” of submission by rational persons – i.e., it is the reasonably *imputed* intention, on the assumption that the person is prudent and rational. This is why the “Obligation, and Liberty of the Subject, is to be derived” in part “from the End of the Institution of Sovereignty; namely, the Peace of the Subjects . . . and their Defence.”<sup>71</sup> One may wonder how conventional obligations grounded in *voluntary* actions could be grounded in *imputed* intentions one may not actually have. Again, the answer is that for Hobbes directed obligations derive not from subjective mental states, but from the interpersonal meanings of utterances and actions – from the *signs* of a person’s will as reasonably interpreted by others on the presumption he is rational.

#### 6.4 The Relation between Natural Law and Civil Law

The foregoing analysis clarifies one of the central puzzles of Hobbes’s philosophy of law, concerning the relationship between natural law and positive or civil law and whether there are antecedent normative constraints on sovereign authority. Hobbes defined “Law in generall” as the command “of him, whose Command is addressed to one formerly obliged to obey him,” and civil law in particular as “those Rules” that

to every Subject . . . the Common-wealth hath Commanded him, by Word, Writing, or other sufficient Sign of the Will, to make use of, for the Distinction of Right, and Wrong.<sup>72</sup>

On the basis of such declarations, Hobbes has been frequently construed as committed to two theses ostensibly making him a precursor to legal positivism: first, that the source – and only source – of law and legal obligation is the sovereign’s positive commands; second, that whether

<sup>71</sup> L 21.10: 336.    <sup>72</sup> L 26.2–3: 414, italics removed.

a precept counts as law does not inherently depend – as it did for classical natural-law theorists – on its conformity to any antecedent natural norms.<sup>73</sup>

As recent critics have pointed out, however, Hobbes did appear to hold that natural laws constrain what the sovereign can command as law:<sup>74</sup> “whatsoever is not against the Law of Nature, may be made Law in the name of them that have the Sovereign power,” which insinuates that no sovereign can enact as civil law anything contrary to natural law.<sup>75</sup> Although natural laws are not properly laws for which anyone is accountable until incorporated into civil law, they do nevertheless articulate precepts that are normative prior to sovereign command. Hobbes also suggested that natural law, beyond placing negative limits, is positively incorporated into the civil law of every commonwealth:

The Lawes of Nature . . . are not [in themselves] properly Lawes, but qualities that dispose men to peace, and to obedience. When a Commonwealth is once settled, then are they actually Lawes, and not before; as being then the commands of the Common-wealth; and therefore also Civill-Lawes . . . The Law of Nature therefore is a part of the Civill Law in all Common-wealths of the world.<sup>76</sup>

The challenge is to understand why Hobbes thought that natural law is incorporated into the civil law of *every* commonwealth. For as Gregory Kavka has pointed out, from the premise that natural laws are *not* laws until incorporated into civil law, it does not follow that the sovereign *does* incorporate them into civil law. A sovereign might pass a natural law over in silence, explicitly disavow it, or, as Kavka notes, “enunciate civil laws that contradict it.”<sup>77</sup> The sovereign may, for example, enshrine as law a command to “Kill all mediators on sight,” in apparent violation of the fifteenth law of nature.<sup>78</sup> Hobbes’s argument works only if the incorporation of natural law were a systematic feature of the structure of civil law as such.

Some commentators have suggested that the resolution to this tension, between the ostensibly positivist and natural-law elements in Hobbes’s philosophy of law, lies in natural law’s putatively “self-effacing” character.

<sup>73</sup> These are the so-called Sources Thesis (in its sovereign command version) and Separation Thesis. The natural norms are typically understood to be moral, but some reject legal positivism on the grounds that there are non-moral, formal, or procedural norms to which laws and their enactment must conform (Fuller 1969). For positivist readings of Hobbes, see Hart (1961: 62–64); Watkins (1973: 114); Hampton (1986: 107–110); Kavka (1986: 249).

<sup>74</sup> Murphy (1995); Dyzenhaus (2001, 2010, 2012); Cuffaro (2011). <sup>75</sup> L 26.40: 448.

<sup>76</sup> L 26.8: 418. <sup>77</sup> Kavka (1986: 249). <sup>78</sup> L 15.29: 236. The example is Murphy’s (1995: 857).

On this reading, while in principle no precept contrary to the laws of nature can be law, the only normative requirement of natural law in a commonwealth is to obey sovereign command. Natural law is therefore intrinsically a part of civil law as its normative foundation, but places no substantive constraints on what can count as civil law.<sup>79</sup> The weakness of readings such as this – which leave no room for the substantive role of natural law in shaping the *content* of the legal order – is that they ignore Hobbes’s characterization of the persistent role played, even after the commonwealth’s establishment, by *specific* laws of nature whose point is to regulate the functioning of the legal system and the role of judges in it. These include natural laws concerning cruelty in punishment (seventh law), equity in judicial interpretation (eleventh), and the character of judges and judicial testimony (sixteenth through eighteenth).<sup>80</sup>

Hobbes’s theory of speaker-meaning and why reasons of the good constrain and supplement reasons of the right clarifies this puzzle, by showing why a *system* of civil law implicitly incorporates the laws of nature and thereby renders subjects accountable for them – even when they are not explicitly legislated.

Civil laws are commands. The question is what the relation is between the sovereign’s commands and his will or intention. On the one hand, because commands are acts of communication, they cannot consist in the commander’s actual will – which is wholly private and so inaccessible to subjects: “the will of another, cannot be understood, but by his own word, or act, or by conjecture taken from his scope and purpose.”<sup>81</sup> Just as covenants are grounded in signs of the will, so too must commands consist in outward *signs*: “the Law is a Command, and a Command consisteth in declaration, or manifestation of the will of him that commandeth, by voyce, writing, or some other sufficient argument.”<sup>82</sup> Commands are by nature publicly accessible: “Commands, are to be signified by sufficient Signs; because a man knows not otherwise how to obey them.”<sup>83</sup> The paradigmatic signs through which the sovereign commands his subjects are, of course, the written words through which law is articulated. The problem is that “the significations of almost all words, are either in themselves, or in the metaphoricall use of them, ambiguous; and may be drawn in argument, to make many senses.”<sup>84</sup> This is one reason why laws “have need of Interpretation.”<sup>85</sup>

<sup>79</sup> Bobbio (1993); Lloyd (2009). On Murphy’s (1995) variant, natural law requires acting only if others will reciprocate, which they will not do if a civil law forbids the action.

<sup>80</sup> L 15.19: 232; 15.23: 236; 15.30–32: 238. On the centrality of these laws, see Dyzenhaus (2001).

<sup>81</sup> L 26.15: 424. <sup>82</sup> L 26.12: 422. <sup>83</sup> L 26.4: 416. <sup>84</sup> L 26.26: 436. <sup>85</sup> L 26.21: 430.

On the other hand, the function of publicly accessible linguistic expressions is to serve as signs of the *speaker's intention*. To understand the law one must understand not just expression-meaning (and so the semantic and syntactic conventions governing linguistic expressions), but speaker-meaning: “no written Law, delivered in few, or many words, can be well understood, without a perfect understanding of the finall causes, for which the Law was made.”<sup>86</sup> The law therefore does not consist in “the Letter” or “whatsoever can be gathered from the bare words” but, rather, in the intention of the sovereign as signified by the utterance of those words: “For it is not the Letter, but the Intendment, or Meaning; that is to say, the authentique Interpretation of the Law (which is the sense of the Legislator,) in which the nature of the Law consisteth.”<sup>87</sup> This is another reason why laws “have need of Interpretation.”

This returns us to the original problem: the sovereign's *actual* will is not directly accessible to subjects. This is why “the sense of the Legislator . . . in which the nature of the Law consisteth” comprises not the sovereign's actual will, but the will *reasonably imputed* to him. Subjects seeking to obey their sovereign's commands must therefore “conjecture” about his intentions to interpret the signs of his will. Understanding speaker-meaning requires not just (a) understanding the semantic and syntactic conventions that govern the “bare words” of the law, but also (b) understanding the pragmatic context in which he uttered or tacitly reiterates them, and (c) making reasonable assumptions about his general dispositions and purposes.<sup>88</sup> One set of such assumptions is that the sovereign is prudent and rational and therefore intends to abide by natural law: “the will of another, cannot be understood, but by his own word, or act, or by conjecture taken from his scope and purpose; which in the person of the Common-wealth, is to be supposed alwaies consonant to Equity and Reason.”<sup>89</sup> If the sovereign commands me to “Kill all mediators on sight,” but I reasonably discern that the fifteenth law of nature prescribes allowing “safe Conduct” to “all men that mediate Peace,”<sup>90</sup> then I have a normative reason to interpret his command and natural law in a way that reconciles the two precepts. I may have an epistemic reason to think, for example, that the sovereign's command has an implicit rider restricting its scope to duplicitous or unauthorized mediators, whose activities are

<sup>86</sup> L 26.21: 430.      <sup>87</sup> L 26.26: 436; 26.20: 428.

<sup>88</sup> EL 5.8; 13.8; 15.6; DCv 17.18. On pragmatics in Hobbes, see Zarka (1999); Abizadeh (2015).

<sup>89</sup> L 26.15: 424.      <sup>90</sup> L 15.29: 236.

noxious to peace, and that the fifteenth law of nature has an implicit rider restricting its scope to mediators authorized by the sovereign.<sup>91</sup>

This already suggests how natural laws shape the content of the civil laws to which subjects are accountable. Since the civil law is constituted not by the “bare words” but by their meaning, which in turn depends on imputing to the sovereign a will to abide by the laws of nature, the civil law must incorporate them. Yet we should tread carefully here, because however reasonably a subject may interpret the publicly accessible signs of the sovereign’s will, no subject is *accountable* to his own private interpretation. He is accountable, and is held accountable, to the commonwealth’s *public* interpretation as authorized by the sovereign: “the Interpretation of all Lawes dependeth on the Authority Sovereign,” and those “Lawyers” who make “the Legislative Power depend on private men, or subordinate Judges” are deeply mistaken about the nature of sovereignty and law.<sup>92</sup> And the *sovereign’s* own interpretation depends on his actual, not reasonably imputed, will. Faced with the command to “Kill all mediators on sight,” subjects might therefore have a strong prudential reason to consider whether the sovereign will hold them accountable to a civil law incompatible with their own interpretation of natural law, however reasonable. If the sovereign interprets and applies the law to particular cases in his own person, then his subjects are held accountable to civil laws shaped by natural laws only if the sovereign, as a matter of contingent fact, incorporates them into his practice of legislation and adjudication. But this could not explain Hobbes’s claim that *every* commonwealth incorporates them.

The systematic incorporation of natural into civil law arises instead from a further consideration. Hobbes took it for granted that in any society large enough to constitute a commonwealth, the sovereign could not adjudicate all legal disputes in his own person. The practice of adjudication, he assumed, requires judicial representatives authorized to act on his behalf. Civil law requires a legal *system* with a set of judicial institutions.<sup>93</sup> This is why his declaration that “the Interpretation of all Lawes dependeth on the Authority Sovereign” is immediately followed by the clarification that “the Interpreters can be none but those, which the Sovereign, (to whom only the Subject oweth obedience) shall appoint.”<sup>94</sup> In practice, subjects will be held accountable by authorized judges whose interpretation of the law depends not on the sovereign’s actual will, but on the will they impute to

<sup>91</sup> Cf. DLDL II.13.7–9.   <sup>92</sup> L 26.20: 428; 26.10: 420.

<sup>93</sup> For the argument that judicial institutions shape the relationship between natural and civil law, see Dyzenhaus (2001, 2010, 2012).

<sup>94</sup> L 26.20: 428.

him. And the will that is reasonable for judges to impute is consonant with natural law: “if the Sovereign . . . make a Judge, The Judge is to take notice, that his Sentence ought to be according to the reason of his Sovereign, which being alwaies understood to be Equity, he is bound to it by the Law of Nature.”<sup>95</sup> Judges do not merely have an epistemic reason to impute such an intention. Because they represent the sovereign and occupy an official position, they also have a strong prudential reason to interpret the sovereign’s commands in ways compatible with the ongoing health of the commonwealth and its institutions, i.e., in accord with those “Fundamentall Law[s] . . . without which the Common-wealth cannot stand.”<sup>96</sup> For example, if “in whatsoever words any of” the essential rights of sovereignty “seem to be granted away” by the sovereign, “yet if the Sovereign Power it selfe be not in direct termes renounced,” then judges are to interpret legislation on the basis that “the Grant is voyd.”<sup>97</sup> Such fundamental laws, which concern legal and political institutions, constitute sovereignty and are partly articulated in those natural laws for which the sovereign is attribution-responsible: for a sovereign to violate these laws in legislation or adjudication is for him to be in “breach of” the “duty” and “Office of the Sovereign,” a breach consisting in the “Negligent government of Princes” whose eventual “Naturall Punishments” are “Rebellion” and “Slaughter.”<sup>98</sup> This is why judges have reasons not only to *constrain* their interpretation of civil laws but also to *supplement* it in light of their interpretation of natural law:

the Intention of the Legislator is alwaies supposed to be Equity: For it were a great contumely for a Judge to think otherwise of the Sovereigne. He ought therefore, if the Word of the Law doe not fully authorise a reasonable Sentence, to supply it with the Law of Nature; or if the case be difficult, to respit Judgement till he have received more ample authority.<sup>99</sup>

This is not to say that subjects are absolutely guaranteed that the civil laws to which they are held accountable will actually fully incorporate the laws of nature. Any sovereign who actually intends to legislate a precept contrary to natural law remains the ultimate court of appeal and may ultimately adjudicate the case in his own person. Even if he does not adjudicate in person, subordinate judges may fail to make reasonable judgements consonant with natural law.<sup>100</sup> The point is not that natural law will shape the application of civil law in each particular case, but that the epistemic and prudential reasons governing the practice of judicial

<sup>95</sup> L 26.14: 424.

<sup>96</sup> L 26.42: 448.

<sup>97</sup> L 18.17: 280.

<sup>98</sup> L 31.40: 572; 30.3: 520.

<sup>99</sup> L 26.26: 436.

<sup>100</sup> L 26.24: 432.

interpretation systematically dispose the legal order as a whole to incorporate natural law. The system of positive law to which subjects are held accountable – whose laws consists not in the “bare words” codified in statute but in the living *practice* of legislation and adjudication – is systematically disposed to incorporate the laws of nature even if the bare words do not.

None of this, however, supports the recently advanced view<sup>101</sup> that Hobbes was a rule-of-law theorist according to whom sovereign authority is intrinsically legal, the scope of sovereign authority is constitutionally restricted by the law of nature, and the sovereign is accountable to subjects, via the judiciary, for conforming to natural laws. It is one thing for the system of civil law through which sovereign authority is ordinarily exercised systematically to incorporate natural law. But it is quite another thing for sovereign authority intrinsically to be restricted to legal authority. Hobbes explicitly repudiated the latter view, insisting that the sovereign is above rather than subject to civil laws.<sup>102</sup> True, the sovereign can pass civil laws enabling subjects to sue him, but he remains above the law:

If a Subject have a controversie with his Sovereigne . . . grounded on a precedent Law; he hath the same Liberty to sue for his right, as if it were against a Subject; and before such Judges, as are appointed by the Sovereign. For seeing the Sovereign demandeth by force of a former Law, and not by vertue of his Power; he declareth thereby, that he requireth no more, than shall appear to be due by that Law. The sute therefore is not contrary to the will of the Sovereign . . . But if he demand, or take any thing by pretence of his Power; there lyeth, in that case, no action of Law: for all that is done by him in Vertue of his Power, is done by the Authority of every subject, and consequently, he that brings an action against the Sovereign, brings it against himselfe.<sup>103</sup>

The sovereign remains above the law in two senses. First, he could, even in virtue of his specifically *legislative* authority, simply repeal “those Lawes that trouble him.”<sup>104</sup> Second, his sovereign authority is not confined to legislative authority: he may “take any thing by pretence of his Power,” and still act with sovereign “Authority.”<sup>105</sup>

<sup>101</sup> Dyzenhaus (2001: 461, 491; 2010: 457). <sup>102</sup> L 26.6: 416; 29.4: 504.

<sup>103</sup> L 21.19: 342. See DCv 6.15; 12.4. <sup>104</sup> L 26.6: 416.

<sup>105</sup> In defending Dyzenhaus’s interpretation, Fox-Decent (2012: 139) is forced to read “Power” in L 21.19 as *legislative power* (contrary to Hobbes’s explicit contrast between what “the Sovereign demandeth by force of a former Law” and “by vertue of his Power”). Fox-Decent’s textual warrant is L 26.5: 416: “the Common-wealth only, praescribes, and commandeth the observation of those rules, which we call Law.” But Hobbes was claiming that it is only the (representer of the) commonwealth that commands us to obey laws, and not, as Fox-Decent thinks, that the only thing the commonwealth commands is to obey laws. True, at L 27.27 Hobbes suggested that when

### 6.5 Two Puzzles about Natural Right and Natural Law

When Hobbes introduced the right of nature in *Elements*, he initially defined it as the “blamelesse Liberty” that “every man” has “to preserve his owne body and lymbes.” He then elaborated on this in three ways. First, he glossed the right to self-preservation not just as a right to survival, but as a right to preserve oneself “both from death and paine.” Again, Hobbes was invoking the broad notion of self-preservation: the right of nature refers to the general disvalue of pain, not just the specific disvalue of death. Second, he asserted that the right to preserve oneself implies a right to take the *means* to self-preservation: “where a man hath right to the *end*, And the end cannot be attained without the *meanes*,” he has a right “to use all meanes and doe whatsoever action is necessary for the preservation of his body.” Third, he clarified that the right of nature is not premised on something actually *being* the means to self-preservation, but on the individual *believing* it is: “every man by right of nature is Judge himselfe of the necessity of the meanes.”<sup>106</sup>

Having provided this initial formulation, however, Hobbes then claimed natural right is equivalent to a right to do anything at all: “Every man by nature hath right to all things that is to say to do, whatsoever hee listeth to whome he listeth.”<sup>107</sup> The question is how these two characterizations could be equivalent: whereas in Hobbes’s initial formulation one’s natural right seems conditional on one’s subjective judgement, in the latter formulation it seems unconditional.<sup>108</sup> The same puzzle appears in the first edition of *De Cive*, which repeats the discussion in *Elements* almost verbatim. Hobbes there again initially characterized natural right as the “liberty which every man hath” to “use all his efforts to preserve and defend his Body, and the Members thereof from death and sorrowes [*doloribus*],” and glossed this as a liberty to undertake whatever actions he judges to be the means for promoting self-preservation. He then again claimed this is equivalent to a “*ius in omnia*.”<sup>109</sup>

Hobbes was clearly aware of this in-principle discrepancy between the two formulations: both *Elements* and *De Cive* explicitly address the problem. Hobbes argued that the two characterizations are in practice coterminous, on the grounds that whatever individuals will to do they necessarily conceive to be good for themselves – and hence conducive to

the sovereign acts contrary to a law, he thereby abrogates it. But this does not mean that in doing so he is acting “only through law,” as Fox-Decent puts it. It means what Hobbes said: the sovereign acts extra-legally, but in doing so changes the law.

<sup>106</sup> EL 14.6–10. <sup>107</sup> EL 14.10. <sup>108</sup> Kavka (1986: 300–302). <sup>109</sup> DCv 1.7–10.

their own preservation in the broad sense that it either prevents death or pain: “For seeing all things he willeth must therefore be good unto him in his owne judgment because he willeth them, and may tend to his preservation some tyme or other; or he may judge so . . . it followeth that all things, may rightly also be done by him.”<sup>110</sup> The argument relies heavily on Hobbes having equated the notion of being conducive to self-preservation to being conducive to avoiding pain; it also relies on the additional assumption that pleasure and pain are fungible. But whatever one makes of the argument’s substantive merits, the *interpretive* puzzle looks to be solved.

The second edition of *De Cive*, however, complicates matters. Hobbes there added that, although individuals have a natural right to do whatever they will, nevertheless “if any man pretend somewhat to tend necessarily to his preservation, which yet he himself doth not confidently believe so, he may offend against the Lawes of nature.”<sup>111</sup> This seems to suggest that the two formulations can indeed come apart in practice, which, if so, would reintroduce our puzzle. Richard Tuck has accordingly argued, on the basis of this 1647 addition, that from this point onwards Hobbes abandoned the notion of an unconditional natural right to everything.<sup>112</sup> Tuck’s suggestion is implausible. In *Leviathan*, again having initially defined the right of nature conditionally, Hobbes again equated it with a “natural Right of every man to every thing.”<sup>113</sup>

Furthermore, the 1647 addition does not imply the conditional and unconditional versions of natural right can come apart. Hobbes was not asserting that individuals who voluntarily do something they believe unnecessary for self-preservation might act contrary to or beyond their *right* of nature; he was asserting they may act contrary to the *law* of nature. The conditional and unconditional formulations still have the same extension with respect to natural right: prior to giving up one’s right to all things, whatever one did one had a natural right to do so. Hobbes’s point here was that, when one acts in bad faith, one can violate the laws of nature, even while in possession of the right to all things.

The question therefore is what kind of normative precept one could be possibly violating if, by nature, one has a right to do anything at all. In other words, our initial puzzle, concerning the compatibility of Hobbes’s two formulations of natural right, resolves into a second puzzle, about the compatibility of natural right with natural law in the state of nature. On the one hand, Hobbes defined natural right as a natural *liberty*

<sup>110</sup> EL 14.10; cf. DCv 1.10.    <sup>111</sup> DCv 1.10.    <sup>112</sup> Tuck (1979: 125).    <sup>113</sup> L 14.1–4: 198.

implying the absence of obligation. On the other hand, Hobbes also asserted that rational agents in the state of nature are already “bound” or “obliged” by natural law. Yet if rational agents are always already by nature obliged by natural law, how can they also be entirely free from obligation? *Leviathan* conspicuously highlights this apparent contradiction: immediately after having defined natural right and natural law, Hobbes asserted that “RIGHT, consisteth in liberty to do, or to forbear; Whereas LAW, determineth, and bindeth to one of them: so that Law, and Right, differ as much, as Obligation, and Liberty; which in one and the same matter are inconsistent.”<sup>114</sup>

We can resolve this puzzle only by recognizing the distinction between reasons of the good of the right – and only by recognizing that natural law intrinsically concerns solely the former. The right of nature consists in the absence of obligations in Hobbes’s strict, juridical sense for which one is accountable to others. By contrast, the laws of nature intrinsically “oblige” only in the deviant, eudaimonistic sense that they articulate reasons of the good, for which one is not intrinsically accountable to others.<sup>115</sup> The apparent puzzle is an artefact of the failure to discern the two dimensions of normativity in Hobbes’s ethics.

<sup>114</sup> L 14.3: 198.    <sup>115</sup> Cf. Olsthoorn (2015: 24–25).

*Rational Agency and Personhood*

The sixteenth chapter of *Leviathan* stages one of the most striking moments in the evolution of Hobbes's political philosophy. The chapter has no equivalent in *Elements* or *De Cive*, and it introduces, for the first time, Hobbes's theory of personhood, authorization, and representation. Hobbes's exposition sets the stage for an important addition to his account of sovereignty covenants. According to *Elements* and *De Cive*, when would-be subjects covenant to establish a commonwealth, they do so by transferring their natural right to act on their own private judgement about the means to self-preservation.<sup>1</sup> But according to *Leviathan*, subjects not only transfer to the sovereign their natural right, they also authorize him to act in their name.<sup>2</sup> Some scholars have argued that Hobbes articulated this new apparatus in order to address fundamental problems in earlier versions of his theory of sovereignty; others argue he was appropriating the language used by parliamentarians during the English civil wars, to discredit their anti-royalist conclusions; others suggest he used it to unite his theory of sovereignty with his much-expanded theology and ecclesiology.<sup>3</sup> Whatever the merits of these explanations, we should note that Hobbes recognized the centrality of *personhood* to his normative philosophy from early on. The category of person appeared in the first statement of Hobbes's political philosophy: [part II](#) of *Elements*, "De Corpore Politico," opens with the distinction between the "persons naturall" that make up a "multitude," and the "person Civill" that constitutes a unified "body politique" established via covenants.<sup>4</sup> *Leviathan's* innovation is not the notion of personhood but a theory of personhood in terms of representation and authorization.

<sup>1</sup> EL 20; DCv 5.    <sup>2</sup> L 17.13: 260.

<sup>3</sup> For the first, see Gauthier (1969: chapters 3–4); Zarka (1999: chapter 6). For the second, see Skinner (2005). For the third, see Garsten (2010: 522); Abizadeh (2017a).

<sup>4</sup> EL 20.1–2. See DCv 5.7–12; 6.1.

We can appreciate the significance of Hobbes's notion of personhood only by recognizing the central role it plays in articulating the two dimensions of normativity underwriting his ethics. Hobbes's distinction between reasons of the good and the right is premised on a distinction between rational agency and personhood. An essential point of his analysis of personhood and its relation to authorization and representation is to articulate the intrinsically artificial and normative basis on which rational agents occupy the second-personal standpoint – the standpoint from which one may acquire reasons for which one is accountable to others. Insofar as rational agents are capable of answering to reasons, they already have or “own” normative reasons: it is appropriate for others to take them to be attribution-responsible for their mental states and actions, and even *pro tanto* appropriate to hold them attribution-responsible via normative advice or criticism. Taking rational agents to be (and holding them) attribution-responsible for their actions is to attribute a kind of narrative coherence to them both internally, in presupposing that their outward actions express reason-responsive attitudes or dispositions constituting their inner *character*, and externally, in taking those actions to represent an outward *persona*. Taking rational agents to be attribution-responsible constitutes them as persons.

However, rational agents are not, merely in virtue of having and being capable of answering to reasons, ipso facto accountable to anyone: they do not thereby owe it to anyone to answer to advice or criticism, and no one has the standing to demand they conform to their reasons or respond to criticism for having failed to do so. To be accountable, they must first have obligations in the strict, juridical sense. On Hobbes's account, one can acquire obligations only if others take one's actions to express reason-responsive dispositions and to represent a persona. Personhood is the inherently artificial construct necessary for acquiring obligations because obligations require two things: that others (a) recognize one as a rational agent who acts in ways that communicate the intention to bind oneself to others; and (b) accept one's communicated intentions as binding. Only in virtue of being a person to others can one occupy the second-personal standpoint: the normative power to acquire obligations is an interpersonal, not subjective, power. The notion of personhood is necessary not only for Hobbes's theory of *authorization*, but also for his account – already present in *Elements* – of *transferring* one's right of nature and of thereby acquiring obligations.

### 7.1 Entities, Voluntary Agents, and Rational Agents

Hobbes's metaphysics of personhood is premised on a distinction between (a) entity or body, (b) intentional or voluntary agent, (c) rational agent or author, and (d) person.<sup>5</sup>

An *entity* is any individuated thing numerically distinct from other things. Hobbes took up the ancient metaphysical questions of identity and difference and of self-identity over time in chapter 11 of *De Corpore*, where his examples of individuated items – including “ship” and “man” (*homo*), but also “commonwealth” (*civitas*) – indicate that the category includes natural *and* artificial bodies.<sup>6</sup>

The question of self-identity over time arises because any particular thing “may at different times be compared with itself.” What makes an individuated thing remain numerically the same thing later in time? Hobbes began by canvassing three classical answers: unity of matter, form, and the thing's aggregate *accidentia* or properties. He then gave counterexamples to each proposal: the first criterion implies that an old man whose body has changed from youth is no longer the same man; the second absurdly implies that a ship whose parts are gradually replaced with qualitatively identical new parts is numerically the same ship as another ship built from the old parts in the same form; and the third implies that a man standing is not the same man sitting. Hobbes then argued that what grounds the numerical *continuity* of some named thing depends on the basis upon which the thing so-named has been *individuated* by observers.<sup>7</sup> Each of the three criteria consequently applies to different types of named objects. For example, the first proposal applies to ‘body’: something is not the same natural body if its matter changes. The second proposal applies to ‘river’: something is called a single river on the basis of its moving matter continually originating in the same source, so that today's water, which originates from the same fountain as yesterday's water, remains the same river, even though its matter is different. This is for Hobbes what it means for a river to be individuated by its form. The third proposal applies to ‘ship’: a thing is named a ship in virtue of its particular matter in a particular form, so that insofar as any of its material or formal properties changes, it is to that degree a new ship.<sup>8</sup>

<sup>5</sup> LN §28; L 16.4: 244; DC 1.3; 2.14; DH 15. In Latin: *ens* or *corpus*, *corpus animatum sentiens*, *corpus animatum rationale*, and *persona*.

<sup>6</sup> DC 11.7; L 22.10: 352. Cf. AW 12.1–6. <sup>7</sup> Cf. Holden (2004: 96–99). <sup>8</sup> DC 11.7.

Any *real* entity has causal standing: it is an “agent” insofar as it has effects, and a “patient” insofar as it is causally affected by other agents.<sup>9</sup> *Qua* agent it is *causally* responsible for its effects, and so may be “blamed” in a non-normative sense – although unless it is a voluntary agent there is no point in holding it responsible: its behaviour or motions can be causally affected, but not via threats or retaliatory sanctions.

A *voluntary agent* is an animate body capable of “*voluntary actions*” mediated by “thought” or “the Imagination” and consequent to “deliberation” and “*Willing*.”<sup>10</sup> It therefore includes not only humans, but other animals too.<sup>11</sup> The category also includes artificial intentional agents, such as an aristocratic sovereign assembly that, in virtue of procedures of collective deliberation and majority-rule decision-making, has an artificial will. Hobbes went so far as to call artificial agents “Artificiall Animal[s].”<sup>12</sup>

What makes a voluntary agent the numerically same agent over time? Focussing on ‘homo’, *De Corpore* answers that although a man’s body will not be the same body when older, he will still be the same man so long as he has the same “form,” i.e., so long as all his actions and thoughts continuously proceed, like a river’s waters, from the same source: “that man will be always the same, whose actions and thoughts proceed all from the same *principio*, namely, that which was in his generation.”<sup>13</sup> What ensures that an assembly remains the same artificial agent over time is not the persistence of its matter (i.e., the natural bodies of the individuals who compose it), but the effective persistence of its constitution (i.e., the collective decision-making rule by which its unified will or intention is formed).

Although humans and other animals are alike in being voluntary agents, they differ in two significant respects: adult humans are normally *rational agents* and *natural persons*. Hobbes made the first point by saying that “the names *Man* and *Rationall*, are of equall extent, comprehending mutually one another.”<sup>14</sup> It is true that, according to Hobbes, animals can deliberate and that some may even be capable of a non-linguistic form of reasoning. But a rational agent is capable of a specifically linguistic, propositional kind of reasoning, which uniquely enables him to reason inferentially from premises, and so “by words reduce the consequences he finds to generall Rules.”<sup>15</sup> It is this linguistic, propositional kind of reasoning that enables rational agents reflectively to answer to reasons. Hence not all natural voluntary agents are rational agents. By contrast, all artificial voluntary

<sup>9</sup> DC 9.1.    <sup>10</sup> L 6.1: 78, 6.53–54: 94.    <sup>11</sup> L 6.1: 78; 6.51: 92; 6.53: 92.

<sup>12</sup> L Introduction.1: 16. The category includes sovereign and non-sovereign assemblies (L 19.1: 284; 22.10: 352).

<sup>13</sup> DC 11.7.    <sup>14</sup> L 4.8: 52. Cf. DC 2.14.    <sup>15</sup> L 5.6: 68.

agents are rational agents: any “Artificiall Animal” must be “an Artificiall Man,” i.e., it must have not just a will, but “an artificiall *Reason* and *Will*”<sup>16</sup> arising from linguistically mediated procedures of collective deliberation, reasoning, decision-making, and, as we shall see, authorization.

As we saw in [Chapter 5](#), because voluntary agents deliberate and undertake voluntary actions, they can be the appropriate target of sanctions aimed at shaping their behaviour: they can be *held* causally responsible. But because non-rational creatures cannot answer to normative reasons, they do not have any such reasons and hence are not *normatively* responsible: even their imprudent actions are not the proper object of normative criticism. By contrast, because rational agents have mental states and actions that answer to reasons, it is appropriate to attribute those states and actions to them and to take them to be and hold them attribution-responsible. Hobbes’s term of art for an agent with normative practical reasons is an “*Owner*,” “*Author*,” or “*Dominus*.” A rational agent “owneth” or is the “*Author*” of his voluntary actions.<sup>17</sup> Being a rational agent is a necessary and sufficient condition for being an author, i.e., for having reasons for which one is attribution-responsible. However, rational agents are not, just in virtue of being attribution-responsible, thereby accountable to anyone for anything; only those who are also persons can become accountable. The question is why personhood is necessary, and rational agency insufficient, to occupy the second-personal standpoint from which relations of obligation and accountability arise.

## 7.2 Persons: Representer or Represented?

Hobbes introduced his initial definition of personhood in the opening two paragraphs of *Leviathan*’s celebrated sixteenth chapter:

A PERSON, is he, *whose words or actions are considered, either as his own, or as representing the words or actions of an other man, or of any other thing to whom they are attributed, whether Truly or by Fiction.* / When they are considered as his owne, then is he called a *Naturall Person*: And when they are considered as representing the words and actions of an other, then is he a *Feigned* or *Artificiall person*.<sup>18</sup>

<sup>16</sup> L Introduction.I: 16.

<sup>17</sup> L 16.4: 244. Pitkin (1967: 19); Green (2015). Hobbes used ‘ownership’ in a proprietary sense as well, but here he was using the term in a normative responsibility sense: to be the author of actions is not to have the normative power to sell them, but to be responsible for them, in the sense that their normative status and consequences redound to him.

<sup>18</sup> L 16.1–2: 244.

Many commentators focus on this initial formulation as if it were a complete statement. It is not: Hobbes immediately supplemented it in two ways, one in each of the two succeeding paragraphs.

The first stems from an inconsistency to which the initial formulation appears to lead: the initial formulation appears to define a person as one who *represents* someone or something, but elsewhere in *Leviathan* Hobbes typically wrote of a person as one who is *represented*. For example, although the commonwealth is not a natural person, it is “an Artificiall Man” with “an artificiall Reason and Will” that is “made *One Person*” artificially in virtue of being “by one man, or one Person, Represented.”<sup>19</sup> The artificial person in question is here equated with the *representee* and not, as with the opening two paragraphs of chapter 16, the *representer*. Hobbes wrote of God in the same way in chapters 41 and 42 – as a person in virtue of being represented<sup>20</sup> – and his formulation in *De Homine* appears to enshrine the point: “*a Person is he to whom the Words & Actions of men, either his own or another’s, are attributed. If his own, the Person is natural; if another’s, Artificial [Fictitia].*”<sup>21</sup>

One way to try to resolve this apparent discrepancy is to take Hobbes to have changed his mind. This is the reading proposed by Quentin Skinner, according to whom Hobbes, early in *Leviathan*, equated persons with representers; quietly changed his mind, later in *Leviathan*, to equate persons with representees; and, finally, explicitly announced the change in *De Homine* in 1658.<sup>22</sup> This interpretation fails on two counts. First, the textual evidence breaks up any linear chronology of Hobbes’s putative intellectual evolution. On the one hand, even (plausibly) supposing that chapters 41–42 of *Leviathan* were written later, the apparent equation of persons with representees already appears in chapter 16 itself, in his characterization of the state as an artificial person. On the other hand, Hobbes reiterated his opening formulation, which apparently equates persons with representers, in two works post-dating *De Homine*, where he defined a person as one who “acteth any thing in his own or another’s name, or by his own or another’s authority.”<sup>23</sup> Second, the hypothesis of a quiet change in *Leviathan*’s later parts is ruled out by chapter 42 itself, because Hobbes there gave a formulation akin to *De Homine*’s by *explicitly referring back to his initial formulation* in chapter 16: “a Person, (as I have shewn before, chapt. 13. [*sic*]) is he that is Represented.”<sup>24</sup> Hobbes did not take himself to have changed his mind: he considered both formulations to

<sup>19</sup> L Intro.1: 16; 16.13: 248; 17.13: 260; 18.18: 280.   <sup>20</sup> L 41.9: 772; 42.3: 776–778.   <sup>21</sup> DH 15.1.

<sup>22</sup> Skinner (1999). For criticism, see Runciman (2000).   <sup>23</sup> AB: 310; LL 16: 245.   <sup>24</sup> L 42.3: 776.

express the same conception of personhood.<sup>25</sup> Skinner himself, recognizing some of the problems with his original reading, has more recently suggested that perhaps Hobbes held *both* views: that both representers and representees are persons.<sup>26</sup> This is on track to Hobbes's view, but if it means the representer is one person and the representee another, it is not viable: it would imply, contrary to Hobbes's view, not only that the representer is an artificial person distinct from the artificial person he represents, but also that someone bearing a natural person is *two* natural persons, the one representing, the other represented.

There is no need to conclude Hobbes was deluded about his own views. There is an explanation ready at hand to resolve the apparent discrepancy, and it appears in chapter 16's third paragraph, where Hobbes redescribed representers more precisely, not as those who *are*, but as those who "bear" a person. Something can "be" a person in two distinct senses: although when speaking loosely one can say, as Hobbes himself did, that something *is* a person, it is more precise to say, as Hobbes also frequently did, that it (1) *bears* or *represents* a person to some audience (*qua* representer) or (2) bears the *name* of and is *represented* via the person to some audience (*qua* representee) (or both). Either of these more precise formulations counts as metaphorically "being" a person because Hobbes modelled his conception of personhood by analogy with theatrical representation.<sup>27</sup> Chapter 16's third paragraph likens a *persona* to the artificial "Mask or Visard" actors wear "on the Stage" – the mask that actors "*Personate, . . . Act, or Represent*" via their words and actions for an audience.<sup>28</sup> Thus in contrast to intentional and even rational agency, which for natural bodies Hobbes characterized in non-relational terms as their intrinsic property, he construed personhood as an extrinsic, constitutively relational, and artificial property: one is never a person as such, but is always a person for and in relation to some audience. The implication is that it is possible to be a person for some but not others.

Each adult human *is* only one natural person, but in two distinct senses of 'is': natural persons simultaneously "bear" or "represent," and are "represented" by, their own person to others.<sup>29</sup> Similarly, in the case of artificial persons, in which the representer is distinct from the representee, the former and latter are both the *same* artificial person, albeit in different

<sup>25</sup> Cf. Matheron (1990). <sup>26</sup> Skinner (2002: 190). <sup>27</sup> Vieira (2009). <sup>28</sup> L 16.3: 244.

<sup>29</sup> Hobbes explicitly treated the metaphorical use of 'is' as meaning *represents* in L 44.11: 966–968. Skinner (2002: 192) erroneously claims that for Hobbes adult servants are not natural persons because their master "has the right to speak and act in the name of his entire family." But from the fact that one can be represented it does not follow that one is not a natural person.

senses of ‘is’. An adult king – a natural person – who represents the state to others is an artificial person in virtue of bearing and representing the state, but he is, *qua* representer, the same artificial person the state is, *qua* representee. The king is the state, just as, on Hobbes’s rendering of the Trinitarian doctrine, Jesus was God in the sense that he bore God’s person. This was Hobbes’s underlying conception throughout, and it fully explains why he took both formulations to express the same conception.

Thus the first way Hobbes’s initial formulation is incomplete is in deploying the imprecise, metaphorical language of *being* a person. Hobbes indicated the looseness of his initial formulation by redescribing representers, in the third paragraph, as those who “beare” a person. Chapter 16 opens by saying that a person “is he . . .,” but strictly speaking Hobbes meant that a person is *borne by he*.

### 7.3 Natural Persons Are Authors

A “*Naturall Person*” exists for an audience only when the same natural agent is simultaneously the actor or *representer* and the *representee* of its voluntary actions. To be a representer, one must be “considered” by an audience to represent somebody’s actions; to be a representee, some actions must be “attributed” to one by an audience. One cannot bear a person unless, as a matter of fact, an audience considers one to represent someone, and attributes one’s voluntary actions to someone. The attitude of taking others to be attribution-responsible, and being prone in cases of violations to have critical blaming attitudes towards them, are necessary for the construction of personhood. Thus even a natural person is – despite the label – a relational and artificial construct.<sup>30</sup>

Some have gone further and claimed that Hobbesian personhood is not merely an artificial but also a *social* construct, such that the audience for whom one is a person must necessarily comprise *other* persons.<sup>31</sup> The kernel of truth here is that one is a person to others only insofar as they recognize one as a person. But nothing in Hobbes’s account rules out a natural rational agent who is a natural person only for himself. As long as the rational agent considers the voluntary actions that are naturally his – in the sense that they result from the will caused by his body’s internal motions – to represent his own self, and attributes those actions to himself and so takes himself to be normatively responsible for them, then he is a natural person to himself. That natural personhood can be wholly private is

<sup>30</sup> Vieira (2009: 82–84). <sup>31</sup> Pettit (2008: 56); Vieira (2009: 87).

important for intrapersonal identity over time: a natural rational agent can be the same natural person (for himself) over time even if others fail or cease to recognize his natural personhood.

While personhood intrinsically depends on belief, the belief that someone represents or is represented by another is insufficient. Personhood is not merely a constructed but also a *normative* status: one must not only be “considered” to represent or be represented by the words and actions, but, as Hobbes put it in the fourth paragraph of chapter 16, one must do so “*by Authority . . . from him whose right it is.*” Personhood requires that one be duly *authorized* by the *owner* of the person’s actions (and not merely considered by the audience) to represent or be represented. Being the owner of actions in Hobbes’s technical sense is distinct from being their natural cause. It is to be their *author*.<sup>32</sup> Being the cause of an action is a *natural* property; being attributed the action is a constructed, *artificial* property dependent on beliefs; but being its author is the *normative* property of being attribution-responsible for it.

An owner may be the author of actions in two ways: (a) he may directly *author* actions, insofar as he both directly wills and causes them and is attribution-responsible for them, and/or (b) he may *authorize* or give “warrant to” actions, by commissioning whoever does directly cause them “to beare his Person” and “act in his name” by representing his person via those actions to an audience.<sup>33</sup> What is distinct about an agent bearing a natural person is that the actions that (1) are naturally his in the voluntary, causal sense are the very same actions that (2) are his both (a) in the constructed, artificial sense that they are attributed to him and (b) in the normative sense of being *both* authored and authorized by him. Because the bearer of a natural person not only *authors* the actions of which he is the voluntary cause, but also *authorizes* his own self to represent his person to an audience, he owns – i.e., is attribution-responsible for – the very actions through which he seeks to represent himself. The actions of someone who bears a natural person are therefore “his” actions in three distinct senses: they are his actions *qua* actor or intentional agent in the voluntary and *natural* sense that they result from the will caused by the internal motions of his natural animate body; they are his actions *qua* representee in the relational, *artificial* sense that they are attributed to him or “considered as his owne” by some audience; and they are his actions *qua* owner in the *normative* sense that he is their author and so is attribution-responsible for them.<sup>34</sup>

<sup>32</sup> L 16.4: 244.    <sup>33</sup> Q: 175/215; L 16.3: 244.    <sup>34</sup> L 16.1–4: 244.

It is this coincidence of natural causation, representation, and normative responsibility that justifies labelling this type of personhood “natural” despite being an artificial construct. When an agent is attribution-responsible for the actions that are naturally his and through which he successfully represents his own person, no information is lost by deploying the loose formulation to say that he *is* a natural person relative to some audience: he is a natural person in both senses.

A person is therefore an artificial construct defined in terms of four roles: (1) the “*Actor*” or “*Representer*,” who bears the person, in virtue of being “considered as representing” it via words and actions that are the actor’s in the voluntary and causal sense; (2) the *representee*, which bears the name of and is “represented” by the person, in virtue of having the actor’s words and actions “attributed” to it; (3) the *audience*, which comprises those for whom the person exists, that is, those who “consider” someone the representer, who “attribute” words and actions to the representee, and to whom the person is represented via these words and actions; and (4) the “*Author*” or “*Owner*” of these words and actions, who bears normative attribution-responsibility for them in virtue of authorizing the representation relation.<sup>35</sup>

Hobbes’s initial formulation, in the opening two paragraphs of chapter 16, focusses exclusively on the first three roles, through which personhood is constituted as an artificial construct; it is not until the fourth paragraph, about authorship, that Hobbes introduced the fourth, normative, attribution-responsibility element.<sup>36</sup> If we fail to take into account the normative element of personhood, then we fail to understand why only adult humans – *natural rational agents* – can be (bear) natural persons. Being a natural rational agent is insufficient for being (bearing) a natural person: one must also be *considered* by some audience to represent oneself. This is what the first three elements of personhood imply. But being considered to represent someone is insufficient for bearing a person because one must also be *duly authorized*. Even if some audience were to “consider” an animal to represent itself via its voluntary actions, or to “attribute” the creature’s voluntary actions to itself, it would not thereby become a natural person. To be a natural person one must also be an

<sup>35</sup> Pitkin (1967) overlooks item 3.

<sup>36</sup> In L 16.1–2, the expressions “considered as his owne” and “considered as representing” denote the *empirical* relation of having some actions be “attributed” by an audience to a representee. Neither denotes the normative property of being the owner/author or of being a *duly authorized* representer. If an audience *believes* you represent someone, then you *represent* him to them; but it does not follow you do so with the authority necessary for *personhood*. See Green (2015: 35).

author, and only rational agents can author and authorize actions and hence be attribution-responsible.<sup>37</sup>

#### 7.4 Artificial Persons: True and By Fiction

Whereas a natural person exists for some audience when the same agent is simultaneously representer, representee, and author, an artificial person arises when the representer and representee are distinct. There are two types of representation. The first is *true representation*: representers “*Truly*” represent others when the representers’ actions are authorized or “*Owned* by those whom they represent.” Here the representer is the “Actor,” the representee the “Author,” so that the representative “Actor acteth by Authority” of the represented.<sup>38</sup> The implication is that only persons capable of authoring and authorizing actions – i.e., only rational agents – can be represented in the first, true sense. A rational agent can either represent himself *qua* natural person or be represented by another *qua* artificial person.

The second type of representation is *representation by fiction*. Representation is “*by Fiction*” when two conditions are met: the representee is not an author and so is incapable of having authorized the representer’s actions itself; but the representer’s actions are nevertheless attributed to it on the authority of a third party with dominion over the representee – someone who has property in or jurisdictional authority over it. Something incapable of being an author can be an artificial person, but only by fiction, and can never also be a natural person.

Hobbes asserted at the beginning of paragraph 9 that practically anything can be represented by fiction: “There are few things, that are incapable of being represented by Fiction.” He then proceeded, in the span of five consecutive paragraphs, to furnish five subcategories of things that can be represented only in this way: (1) inanimate things, such as buildings or bridges, (2) animate but irrational things, such as “Children, Fooles, and Mad-men,” (3) imaginary things or “meer Figment[s] of the brain,” such as the pagan gods, (4) the “true God” of Christianity, and (5) a “multitude of men” joined together, such as a unified commonwealth.<sup>39</sup> Hobbes was explicit about the external source of authority in four of his subcategories: “things Inanimate, cannot be authors,” animate but irrational things “can

<sup>37</sup> This suggests Hobbes’s initial formulation is misleading: his distinction between natural and artificial persons turns on the natural person’s self-authorization, not self-representation.

<sup>38</sup> L 16.4: 244.   <sup>39</sup> L 16.9–13: 246–248.

be no Authors,” imaginary things or “Idols cannot be Authors,” and because “the Multitude naturally is not *One*, but *Many*; they cannot be understood for one; but many Authors,” so that a multitude’s representer is authorized not by the unified multitude itself but individually by each natural person “in particular” who composes it. Hobbes spared the Christian God the indignity of being explicitly called incapable of authorship, but he let the fact be understood by listing him in the section on representation by fiction. As a result, Hobbes asserted, the Christian God is himself incapable of authoring any covenants and, like pagan gods, can undertake covenants only via the intermediary of a representer. A bridge, a child, a pagan god, the Christian God, and the state are all examples of an artificial person by fiction.<sup>40</sup> Being fictional here does not negate the normative status of personhood: the actions of a person by fiction are still “owned” or authored; the fiction consists in the fact that the person who owns and hence is responsible for the artificial person’s actions is *another* person, namely the one who duly authorized the representation by fiction. Thus if a child’s master authorizes a “Guardian” to bear the child’s person, then, as long as the guardian acts within the confines of his commissioned authority, the master, not the child or guardian, is responsible for the actions the guardian undertakes *qua* representer, even though the actions are attributed to the child *qua* artificial person.<sup>41</sup>

The relationship in representation by fiction between authorization, representation, and personhood is best illustrated by the case of the commonwealth. The state is an artificial person that a sovereign represents by fiction. It is not the state itself but the individuals who covenant to establish it, who authorize the sovereign to represent the state, “Every man giving their common Representer, Authority from himselfe in particular.” The sovereign bears not only the person of the state, but also the person of each individual covenanter: each authorizes the sovereign to act in the name of both the commonwealth as a whole and in his own name as subject.<sup>42</sup>

Although Hobbes expressed a marked preference for monarchy, he recognized that in principle the role of sovereign representer can be

<sup>40</sup> L 14.23: 210; 18.3: 266. On covenants with God, see Curley (2004). On God as an artificial person by fiction, see Abizadeh (2017a).

<sup>41</sup> L 16.10: 248; 16.5–6: 246.

<sup>42</sup> L 16.14: 250. Pace Runciman (2009). Cf. Martinich (2016: 332). The sovereign is “the Person representative of all and everyone of the Multitude,” to whom individuals have given the authority “to represent them every one,” and whom they authorize in order for “every man to have his person represented.” L 19.1: 284; 19.3: 286; see 28.6: 484.

performed by an assembly.<sup>43</sup> A sovereign assembly instantiates perhaps the most complicated case of representation by fiction. The case is politically significant for Hobbes because he held that a sovereign democratic assembly must lie at the origin of a commonwealth by institution.<sup>44</sup> The case is complicated because, although the role of representer must always be performed by a person, here it is not performed by a natural person: a sovereign assembly can act, and hence represent another, only in virtue of itself being constituted as an artificial person.

How can an assembly be constituted as an artificial person? Only in virtue of an effective collective decision-making rule such as majority voting, because to be a voluntary agent, the assembly requires a unified will.<sup>45</sup> The artificial person of the assembly *qua* represented body is borne by the assembly *qua* representing body: in a sense it is the rule – and the actions of the natural persons guided by it – that represents the assembly and constitutes its members as an artificial person.<sup>46</sup> Furthermore, the assembly *qua* representer represents four sets of artificial persons: it represents itself *qua* represented body; it represents each individual assembly member; it represents the commonwealth; and it represents each individual subject.<sup>47</sup>

If only rational agents can be natural persons, and only rational agents can be “truly” represented artificial persons, this is because only rational agents possess the normative property of being an author. Conversely, even if a natural rational agent is not “considered” by others to be a natural person, he is still an author: natural rational agents bear attribution-responsibility for their actions even if they lack the artificial elements (1–3) necessary for natural personhood. Being *considered* to represent oneself and being considered to be represented by one’s own actions, and so being taken to be attribution-responsible, are necessary for being a natural person; but neither condition is necessary for being (4) the author of, and hence attribution-responsible for, one’s own actions. By contrast, an artificial rational agent cannot be a voluntary agent at all without also being considered by an audience to be a person. The audience must at least include the persons who compose the representative assembly and who have authorized and are guided by the collective decision-making procedure. Artificial agents are all rational agents *and* artificial persons.

<sup>43</sup> L 19.1: 284. <sup>44</sup> Abizadeh (2016).

<sup>45</sup> On the role of majority rule in constituting a mere multitude as an assembly, see EL 20.3; L 18.1: 264.

<sup>46</sup> This suggests Hobbes was closer to Hart (1961: 62) – who argued legal validity requires a secondary rule of recognition – than Hart recognized.

<sup>47</sup> For the first and third, see L 23.2: 376.

### 7.5 Accountability Is Interpersonal: No Accountability to Oneself

Whereas rational agents can be attribution-responsible without being persons, no one can be accountable without being a person. On Hobbes's account (1) the obligations for which one is accountable can be acquired only if one is taken to be attribution-responsible and hence considered a person. An audience can take agents to be attribution-responsible for their actions only if it attributes to them the rationally reflective attitude that their actions, and the attitudes they express, are justified.<sup>48</sup> This is why only rational agents can be attribution-responsible and why holding people attribution-responsible by advising or criticizing them targets their reflective, rational capacities.

Neither being considered attribution-responsible nor being a person is an intrinsically interpersonal relation, because one can take oneself to be attribution-responsible. Accountability, by contrast, is both intrinsically and genetically *interpersonal*. It is intrinsically interpersonal because according to Hobbes (2) one can be obligated only to *other* persons. It is genetically interpersonal because (3) obligations can be acquired only if other persons take one's actions to be the sign of an intentional offer normatively to bind oneself to them and they accept that offer as binding. One can *be* accountable to others only if they *take* one to be accountable to them. The normative power to acquire obligations is intrinsically interpersonal.

The fundamental purpose of the category of personhood in Hobbes's ethics is to furnish the position from which rational agents are able to occupy the second-personal standpoint. This is why, after fleshing out what personhood is in paragraphs one through four in chapter 16, Hobbes immediately turned, in the fifth through eighth paragraphs, to the implications of his account for covenants and directed obligations. Focussing on artificial personhood, where the representer is distinct from the author, Hobbes declared that when the "Actor maketh a Covenant by Authority" or "Commission" of the author, "he bindeth thereby the Author, no lesse than if he [the author] had made it himselfe."<sup>49</sup> Hobbes had already declared, in chapter 14, that to become "bound" is to acquire a directed obligation for which one is accountable to others.<sup>50</sup> His point here was that if an author duly authorizes another to be a representer, then the author can artificially acquire rights and obligations through the representer's voluntary actions. But the author can do so only if his act of authorization

<sup>48</sup> Pettit (2008: 58–59).    <sup>49</sup> L 16.5: 246.    <sup>50</sup> L 14.7: 200.

is socially *effective*, i.e., only if other persons “consider” the representer to represent the representee. Artificial personhood is necessary for covenanting in this way. The case of artificial personhood, in turn, is supposed to be on a par with “the nature of Covenants between man and man in their naturall capacity,”<sup>51</sup> i.e., *qua* natural persons. Only those who consider each other to be persons are able to contract with each other. This is why personhood is a fundamental part not only of Hobbes’s account of how the sovereign is *authorized* to bear the person of the commonwealth, but also of how in covenanting to establish sovereignty individuals *transfer* their right of nature to him. Hobbesian personhood is not itself intrinsically social, but it makes occupying the second-personal standpoint possible.

Hobbes’s thesis that one can be attribution-responsible for one’s actions independently of interpersonal relations, but cannot be obligated and thus accountable except interpersonally, reinforces the division of his ethics into two distinct, mutually irreducible dimensions of normativity. We can see this by contrasting Hobbes’s view to Christine Korsgaard’s. She rejects the division of normativity into two distinct dimensions on the grounds that (a) *all* normative practical reasons are ones for which we are accountable, because these arise only from self-legislation and, in order successfully to give oneself laws, one must first be accountable to oneself for conforming to them. Indeed, she argues that (b) second-personal accountability to others presupposes the prior and more fundamental relation of accountability to oneself.<sup>52</sup>

On Hobbes’s view, by contrast, one can take oneself to have reasons from a first-personal perspective, and so respond to them, without presupposing that one owes it to oneself to do so. This is because the relation of owing or accountability reduces to the owed party having the standing to demand compliance and, in cases of failure, to react by reiterating and vindicating the initial obligation in ways that demand a normative response. For example, Hobbes thought people typically fail to conform to the laws of nature because of “Erroneous Inferences.”<sup>53</sup> People making such errors in reasoning may feel disappointed with and regret their actions’ painful outcomes, but nothing in Hobbes suggests any grounds for people to flagellate themselves for their mistake or to feel reactive passions such as guilt.<sup>54</sup> It is one thing to take something to be a reason, quite another to take it to be a reason grounded in an authoritatively binding command or law that one’s past self has given to one’s subsequent

<sup>51</sup> L 16.5: 246.    <sup>52</sup> Korsgaard (1996b, 2007).    <sup>53</sup> L 27.10–12: 458–460.

<sup>54</sup> Except insofar as they take natural laws to be *divine* commands.

selves. And Hobbes would have rejected Korsgaard's second claim (b) precisely because his ethics rules out acquiring obligations owed either to no one or directly to oneself. Persons can bind their future selves to their present self only mediately and virtually, according to Hobbes, by binding themselves to some other person whom they authorize to be their representer.<sup>55</sup>

Thus, unlike Korsgaard, who grounds *inter*personal obligation in an agent's authority over herself and her consequent ability to bind herself *intra*personally over time, Hobbes took interpersonal obligations to be basic, and ruled out intrinsically private, intrapersonal accountability. One's present self is not accountable *to* one's past self; nor is one's present self accountable *for* one's past self's actions unless at some point one acquired a directed obligation to someone else.

Two premises undergird Hobbes's view. The first stems from his metaphysics of personhood and time, according to which only one's *present* self can be a natural person: one's past self, like all "things *Past*," has no "being in Nature," but is "a being in the Memory onely."<sup>56</sup> One's past self can presently be at most an artificial person. And the duly authorized representer of one's past self is paradigmatically (albeit not necessarily exclusively) one's present self. This furnishes Hobbes's first premise: even if one could be obligated to one's past self, one could nonetheless release oneself from the putative obligation *at will* because one's present, supposedly obligated, person also represents the putative obliger! Hobbes's second premise is that to be under a putative obligation one could release at will is to be under no obligation at all. It follows that there is no such thing as purely intrapersonal obligation. It is not "possible for any person to be bound to himselfe; because he that can bind, can release; and therefore he that is bound to himselfe onely, is not bound," because "he is free, that can be free when he will."<sup>57</sup>

G. A. Cohen has argued that the second premise of Hobbes's argument rests on a "big mistake." In particular, "Hobbes is wrong that, if you can free yourself at will, then you are already free" because it is wrong to suppose "that if I *can* repeal" a putative obligation it thereby "fails to bind me even when I have not *yet* repealed it."<sup>58</sup> Cohen's argument fails to account for the logical and temporal interstice between a person's intention or will and actions. Recall that obligations, in Hobbes's strict, juridical

<sup>55</sup> On virtual self-binding, see Abizadeh (2013).

<sup>56</sup> L 3.7: 42. For the metaphysics of time, see DC 7. <sup>57</sup> L 26.6: 416. See DCv 6.14; 12.4.

<sup>58</sup> Cohen (1996: 170).

sense, furnish reasons *for action* only. Obligations therefore bind one's actions *in foro externo*, but strictly speaking they cannot bind, in the directed sense, one's will *in foro interno*: having a reason to perform an action for which one is accountable can furnish a reason to intend or will its performance, but this reason cannot, on Hobbes's view, in itself be owed to anyone.<sup>59</sup> So in the first place a contract provides a directed reason (obligation) for action. Granting the normativity of Instrumental Transmission, it provides a non-directed reason to desire or will the action's performance if (and only if) desiring or willing to undertake it is a relevant means for performing it.

But the reason to desire or will the action is non-directed, derivative, and conditional: one has it only so long as one shall be obligated to perform the action at the moment of action. Now recall Hobbes's assumption that one can abrogate a putative self-obligation via *willing* its violation – a person can abrogate it *at will*, “when he will.” Cohen grants this, but argues that *until* one has so willed, one is obligated. But since the moment of willing the violation stands logically and temporally prior to acting, the putative obligation to act will have already been abrogated, prior to the moment of acting. Cohen's objection misses Hobbes's assumption that if by the very fact of *willing* to do something one can thereby release oneself from the putative obligation not *to do* it – if the very intention or will to violate constitutes an abrogation – then the putative obligation cannot *ever* bind one's actions and so is no obligation at all. Obligation in Hobbes's strict sense implies accountability, but there is no circumstance under which one could violate, and so be accountable for violating, a putative self-obligation. Without the possibility of accountability, no obligation can exist.

This reply to Cohen is fairly robust: Hobbes would have been entitled to it even if he had supposed that one could be accountable to others for unexpressed mental states. Even if Hobbes had allowed that one could acquire, for example via contract, a directed obligation to *will* certain actions, the obligation to will the action would still be derivative from and conditional on the obligation to perform it. More particularly, the obligation to will would be contingent on it being the relevant means to perform an action one shall be obligated to perform *at the moment of action*. But if one can release oneself of the obligation to act *at will* – if willing to violate a putative obligation to act *itself* abrogates the

<sup>59</sup> See EL 19.7. One may of course owe it to others to *express* the will to fulfil the contract (EL 15.18; L 14.25: 212).

obligation – then one shall not be obligated to act at the moment of action. And if one shall not be obligated at the moment of action, then one is not obligated not to will the action. Once again, Cohen's objection would fail. No possible violation, no accountability; no accountability, no obligation.

## *Conclusion*

### *Naturalism and Normativity*

We are now in a position to address the relation between Hobbes's materialism and ethics. I have argued that, according to Hobbes, natural persons have natural, artificial, and normative properties. I shall now argue that Hobbes's materialist metaphysics had two important implications for his analysis of artificial and normative properties. First, his materialism committed him to a reductive analysis of non-normative artificial properties, according to which such properties can be reductively analyzed and re-described in terms of natural properties – ultimately, extension and motion. An object's non-normative artificial properties are real properties of the object (they are properties with causal standing), but they are really the properties of agents who conceive it (they are grounded in the intrinsic properties of observers, not the object). Second, Hobbes's materialism committed him to denying that normative properties are real properties: they do not have causal standing. They are therefore not reducible to extension and motion. Nevertheless, Hobbes construed these “irreal” properties, even when artificial, as the object of truth-apt and epistemically objective propositions. Moreover, the unbridgeable chasm between the two dimensions of normativity partly consists in the fact that all normative properties grounded in reasons of the right, but not all normative properties grounded in reasons of the good, are constitutively artificial. Hobbes was an ethical naturalist of a rather sophisticated kind: normative properties are all reducible to (normative) natural properties, but they are irreal and irreducibly normative and are the object of potentially veridical normative judgements and true normative propositions.

#### **Natural versus Artificial Properties**

Consider the beautiful, fragrant, red flower in my back yard. I am asking you to consider a flower that *actually* exists, not one you might imagine could have possibly existed but does not. First, on Hobbes's account, this

flower is a *natural* body, i.e., it exists independently of any intentional agent's mental representation of it. It is not an *artificial* body like a kingdom or commonwealth, the existence of which depends on being represented in agents' minds. Nor is it a *supernatural* body, something inconceivable to material agents and the existence of which would depend on being represented in the mind of a divine agent not subject to laws of causation. Hobbes defined a natural body as a spatially extended entity that "dependeth not on the imagination, but is a reall part of . . . the *Vniverse*."<sup>1</sup> Second, unlike my sensory perception of the flower, the flower's existence is *ontologically objective*: the flower would exist even if there were no subjective mental states in the world.<sup>2</sup> Third, in virtue of being a natural body the flower is also a *real* entity: "the *Vniverse*, being the Aggregate of all Bodies, there is no reall part thereof that is not also *Body*; nor any thing properly a *Body*, that is not also part of (that Aggregate of all *Bodies*) the *Vniverse*." To say an entity is real is to say that, at least partly in virtue of its intrinsic properties, it has causal standing in the world: my flower, for example, shields the soil underneath from direct sunlight. (A real artificial body, such as a "reall, not a metaphoricall Kingdome,"<sup>3</sup> has causal standing in virtue of the natural bodies composing it – bodies who possess a common representer forging in them "a reall Unitie"<sup>4</sup> that disposes them to act from a unified will.)

Hobbes's metaphysical naturalism consists in his view that all real properties are reducible to natural properties. Constitutively artificial properties, which obtain only in virtue of being represented by intentional agents as obtaining, are reducible to the real, ontologically subjective, natural properties of intentional agents. The *real*, artificial properties of objects are grounded not in the objects' but in observers' intrinsic properties: they are not *really* the property of the objects, but of observers. And supernatural properties, which purportedly obtain in virtue of being represented in the mind of an immaterial agent, are not real. Hobbes's reductionist corpuscularianism, in turn, consists in the view that natural bodies have only two sets of necessary intrinsic properties in terms of which all real properties may be reductively analyzed: extension or magnitude (and hence shape) and motion.<sup>5</sup>

Some of my flower's natural properties are *intrinsic* to it; for example, it has a certain shape and height. But some of its natural properties are

<sup>1</sup> L 34.2: 610.

<sup>2</sup> 'Objective' in our contemporary sense meaning mind-independent, not in the scholastic sense meaning objects of mental representation (subjective in our sense) (Nuchelmans 1983: chapter 1).

<sup>3</sup> L 35.11: 642.    <sup>4</sup> L 17.13: 260.    <sup>5</sup> LL 9: 125.

*extrinsic*, existing only in relation to other bodies; for example, the flower is liable to be crushed by large falling stones. Both sets of properties are real insofar as they have causal standing; and they are natural in the constitutive sense insofar as they do not depend on intentional agents representing the flower as having those properties. But these constitutively natural properties may turn out to be *genetically* artificial: my flower's shape and size, for example, are the causal effect of my son's horticultural diligence. And the flower's constitutively natural properties are not only real, they are really properties *of the flower*: their causal standing depends, at least in part, on the flower's intrinsic properties. Furthermore, all of the flower's natural properties are ontologically objective. (By contrast, some of the real, natural properties of the squirrels who frequent my backyard – such as having mental states like pain – are ontologically subjective.) Finally, that my flower has these natural properties is an *epistemically objective* fact: the truth of the proposition that my flower is thirty centimetres tall does not depend on anyone's subjective attitudes, such as their conceiving it to be thirty centimetres or assenting to the proposition that it is.

On Hobbes's account, however, the sensory quality of being red – the flower's colour as it appears to regular observers like me – is not a real property of the flower. As Hobbes put it in *Elements*, “whatsoever Accidents or qualities our Senses make us thinke there be in the world, they are not there, but are Seeminges and Apparitions only. The things that really are in the world without us are those motions by which these Seeminges are caused.”<sup>6</sup> To be sure, that regular observers *see* the flower as red is a real property: it is part of the causal explanation for why our beloved flower will be picked by strangers. But being red is one thing, being seen as red another. That the flower is seen as red is a real, artificial property of the flower, but it is really the property *of observers*: it wholly reduces to a natural, ontologically subjective property of theirs. (More precisely – given that most regular observers are not red either – it is an ontologically subjective property of the content of regular observers' mental states.<sup>7</sup>)

One of my flower's real properties is (not that it is red but) that it normally causes regular observers to see it as red. This extrinsic, dispositional property is really a property of the flower because the flower's intrinsic properties constitute it: the flower's extension and/or motions do not constitute the entirety but are a constituent part of the dispositional

<sup>6</sup> EL 2.10. Cf. L 45.14: 1030.

<sup>7</sup> Mental states, albeit ontologically subjective, are for Hobbes real and natural properties of the agent's brain (L 34.17: 622).

property. The flower's dispositional property is also a natural property: although we cannot specify the property without referring to the flower as the intentional object of agents' mental states, the property itself does not consist in or depend on there *being* such intentional states. All minds in the world could be obliterated, the flower would still really have the dispositional property: the property is ontologically objective. It is also an epistemically objective fact that the flower possesses this property: the flower would be disposed to cause regular observers to see it as red even if they failed to attribute *that* property to it (and believed instead that a demon causes them to see it as red). Finally, the property is an actual property of the flower – as opposed to the merely possible property of causing regular observers to see it as blue.

If the term 'red' denotes the sensory quality regular observers perceive in the flower, then the proposition 'The flower is red' is always false on Hobbes's account; but if the meaning of 'red' is settled scientifically, according to the reforming definition proposed by Hobbes, so that it denotes the flower's dispositional property, then 'The flower is red' is true of my flower. This dispositional property is a real, extrinsic, natural, ontologically objective, epistemically objective, and actual property of the flower.

Contrast the flower with the man in *Don Quixote* who fancied himself a knight. On Hobbes's account, Don Quixote is not a real object, for, unlike my flower or a real kingdom, he is not composed of any natural body: his "Existence" depends entirely on the "fancy of man."<sup>8</sup> He is an *irreal*<sup>9</sup> object – a purely fictional character in Cervantes's novel. There are agents who conceive or "consider" him, so he is an intentional object, but he is *merely* an intentional object: Don Quixote is a "meer Figment of the brain." Such "Idols of the brain, [are] nothing reall," and insofar as Don Quixote is purely imaginary he is nothing more than "a Fiction of the mind."<sup>10</sup> Don Quixote is not only constitutively fictional and hence artificial, he is also genetically artificial: according to Hobbes, purely fictional objects are created by "poets" who imagine them to possess properties the poets have previously conceived in sundry sensory objects, which properties they compound together in the imagination. This is why "a man can fancy Shapes he never saw; making up a Figure out of the parts of divers creatures; as the Poets make their Centaures, Chimaeras, and other Monsters never seen."<sup>11</sup>

<sup>8</sup> L 34.15: 620.    <sup>9</sup> Skorupski (2010).

<sup>10</sup> L 2.4: 30; cf. 34.3: 612. See the contrast between reality and fiction in L 27.1: 452.

<sup>11</sup> L 45.16: 1032.

Yet although Don Quixote is an irreal, purely fictional object, he is nevertheless, like my red flower, an actual object – unlike other characters Cervantes could have created but did not. He is an actual object insofar as there are agents in our world, me for instance, who do conceive him: he is an intentional object in our world. For Hobbes to say a poet has “created” a fictional object is for him to say the poet has made a merely possible object actual. Don Quixote even has properties; for example, he is mad. But madness is an irreal property of Don Quixote: the real property is that Don Quixote is represented as mad by the relevant intentional agents. And this latter property is not really a property of Don Quixote but of the relevant agents: all truths concerning Don Quixote are reducible to truths about real agents’ mental states and, like the red flower’s artificial properties, all of his properties are really properties of intentional agents. But madness, unlike sanity, is an actual, not merely possible, property of Don Quixote: *qua* intentional object, Don Quixote is considered by the relevant agents to be mad. The relevant agents may even posit irreal, imaginary causes for his madness: Hobbes, for example, thought “the Gallant madnes of Don Quixotte is nothinge else but an expression of such height of vaine Glory, as readinge of Romants may produce in pusillanimous men.”<sup>12</sup> Whether or not, for any particular intentional agent, it is an epistemically objective truth that Don Quixote is mad depends on whether ‘Don Quixote’ means Don-Quixote-as-conceived-by-Cervantes-in-writing-his-novel, or whether it means Don-Quixote-as-conceived-by-another-intentional-agent. If the former, then for anyone other than Cervantes the truth is epistemically objective; if the latter, it is epistemically subjective for the other agent. Don Quixote is an irreal, entirely observer-relative or artificial, ontologically subjective, actual object; and madness is an irreal, observer-relative or artificial, ontologically subjective, actual property of Don Quixote.

We not only create purely fictional, irreal objects, we also project purely fictional properties onto real objects. We might, for example, take delight in imagining ourselves more powerful than we really are and indulge in “a compound imagination” that is “properly but a Fiction of the mind,” such as “when a man compoundeth the image of his own person, with the image of the actions of an other man; as when a man imagins himselfe a *Hercules*, or an *Alexander*.”<sup>13</sup> Such representations are typically impotent: we do not become Hercules by imagining that we are. Even in cases in which one is fully aware that one is indulging one’s imaginations, Hobbes did not take

<sup>12</sup> EL 10.9.    <sup>13</sup> L 2.4: 28–30.

kindly to this phenomenon. He characterized self-aware, fictional self-indulgence as a species of “Vain-glory,” and subjected it to derision: “Vain-glorious men, such as without being conscious to themselves of great sufficiency, delight in supposing themselves gallant men, are enclined onely to ostentation; but not to attempt: Because when danger or difficulty appears, they look for nothing but to have their insufficiency discovered.”<sup>14</sup> Self-indulgent vainglory is unworthy of a mature character: “the feigning or supposing of abilities in our selves, which we know are not, is most incident to young men, and nourished by the Histories, or Fictions of Gallant Persons; and is corrected often times by Age, and Employment.”<sup>15</sup> Despite casting these aspersions, however, Hobbes conceded that pleasurable fictions are an inevitable feature of mental life – so much so that even self-indulgent vainglory about one’s own powers is not blameworthy: “For to be pleased in the fiction of that, which would please a man if it were reall, is a Passion so adhaerent to the Nature both of man, and every other living creature, as to make it a Sinne, were to make Sinne of being a man.”<sup>16</sup>

Just as the self-indulgently vainglorious man is aware he is not really a hero, most people who think about Don Quixote realize he is a purely fictional object. But sometimes people mistake purely fictional objects and properties for real ones: call these *putatively real* fictional objects and properties. Some vainglorious people – like Don Quixote – mistake the powers they imagine themselves to possess for real powers. This second, self-delusional form of vainglory is also typically impotent – believing we are Hercules does not make it so. Hobbes took a rather dimmer view of self-delusional vainglory: on Hobbes’s telling, it is either humiliating or dangerous. Those suffering from it “are enclined to rash engaging; and in the approach of danger, or difficulty, to retire if they can: because not seeing the way of safety, they will rather hazard their honour, which may be salved with an excuse; than their lives, for which no salve is sufficient.”<sup>17</sup>

Putatively real fictions exercised Hobbes above all because he took them to lie at the root of religion. Ghosts, demons, and pagan gods are his paradigmatic examples of putatively real fictional objects – which, like objects appearing “in a Dream, to one that sleepeth; or in a Looking-glasse, to one that is awake,” are mere apparitions; but “men not knowing that such apparitions are nothing else but creatures of the Fancy, think [them] to be reall, and externall Substances.”<sup>18</sup> His examples of putatively real, fictional properties projected onto real objects include colours, but also cases in which some “Wood, Clay, or Metall” artefact, such as the

<sup>14</sup> L II.11: 154.    <sup>15</sup> L 6.41: 88.    <sup>16</sup> L 27.1: 452.    <sup>17</sup> L II.12: 156.    <sup>18</sup> L 12.7: 166.

genetically artificial but really existing “Golden Calfe” crafted by Aaron, is taken to be a god.<sup>19</sup> The political problem with putatively real fictions is that subversive opinion leaders – clerics above all – can peddle them to manipulate people’s fears and sew rebellion. This is because of two related ways in which fictions can become potent.

The purely fictional but putatively real properties of gods – such as the terrifying supernatural power to damn people for eternity – correspond to the ontologically subjective but real properties of intentional agents, such as the delusional belief in the gods’ damning power. The danger of these latter properties stems from their causal standing.<sup>20</sup> The gods’ unreal but *actual* properties correspond to observers’ ontologically subjective but *real* properties. Our belief that the idol can damn us to hell does not give it the power to do so – our representation is not directly potent – but it can be *indirectly potent*. For the idol acquires a real property in virtue of our representation: our representation makes the idol fearsome. It disposes us to fear the idol, and this is no fiction. To be sure, this is not really the idol’s property – it is not grounded in its intrinsic properties – but it is a real property.

Moreover, sometimes representations are *directly potent*: objects end up acquiring real properties that artificially simulate the very properties we project onto them – precisely in virtue of our projection. If we represent someone as powerful, we may, in virtue of our representation, make him powerful: “Reputation of power, is Power.”<sup>21</sup> Sometimes, our representation depends on delusion: our representation of the idol as powerful may depend on our representation of it as supernatural. But other times it may not: our representation of a commander as powerful may depend on the fact he is reputed to be powerful, even as our representation of him as powerful itself constitutes his reputation.

Consider “artificial signs” such as the word ‘rain’, which, on Hobbes’s account, is a sign used by English speakers to signify their conception of rain. Like the metal idol or the commander, the word ‘rain’ is a real entity consisting in a body, such as ink markings on paper or moving vocal cords, which produce a visual or aural image, which in turn causes observers or hearers to conceive of rain. This is a real, dispositional property of ‘rain’ for Anglophones and, just like the idol’s fearsomeness or commander’s power,<sup>22</sup> it is an artificial property. It is a constitutively artificial property of ‘rain’ insofar as it depends on the word – not necessarily each token but the type – being represented by language users as having that property.

<sup>19</sup> L 45.16: 1032; 45.23: 1036.    <sup>20</sup> Abizadeh (2013).    <sup>21</sup> L 10.5: 132.    <sup>22</sup> Cf. L 10.1–2: 132.

Because the property is not grounded in any of the word's natural properties, it is a wholly artificial property of 'rain' – which is why 'pluie' could have served just as well. The fact that 'rain' is a sign of conceptions of rain is therefore really a property of English speakers – an ontologically subjective but epistemically objective property of theirs. Hobbesian “natural signs” are artificial too in this *constitutive* sense: thick clouds are a natural sign of rain only insofar as *observers* take them to be evidence for the (imminent) presence of rain. Natural signs do not “as signs, promise any thing . . . at all, but we from them; nor do the clouds, but we, from seeing the clouds, say it shall rain.”<sup>23</sup> The reason Hobbes called them *natural* signs is because, although they are signs only insofar as observers interpret them as such, the interpretation is based on observing a correlation between sign and significate independent of convention: a natural sign is *genetically* natural. In contrast to clouds, however, the word 'rain' is not just a constitutively but also a genetically artificial sign of rain: the relation between word and significate arises thanks to voluntary convention, whether express or tacit.<sup>24</sup>

The difference between the constitutively artificial properties of linguistic signs and those of idols is that the wholly artificial properties of words like 'rain' do not depend on *erroneous* beliefs about their natural properties: 'rain' can function as a sign of conceptions of rain without anyone being under the delusion that this is a natural (or supernatural) property. Some speakers – such as the Adamic essentialists ridiculed by Hobbes – might mistakenly think linguistic properties are natural or supernatural.<sup>25</sup> But linguistic properties do not depend on such illusions.

Linguistic properties depend instead, according to Hobbes, on the directly potent decision by language users to take some word to function as a sign of some conception(s). This deliberate, artificial “imposition” of meaning, as Hobbes put it, consists in assigning what John Searle has called a *status function* on words:<sup>26</sup> the status function of 'rain' consists in the constitutive rule, applied by English speakers, that in the appropriate syntactic and pragmatic contexts 'rain' counts as a sign of a conception of rain. On Hobbes's account, serving a particular function cannot be an ontologically objective property, because only mental states have final causes or ends (which Hobbes took to be analytically reducible to efficient causes): there are no ontologically objective purposes in nature.<sup>27</sup> That an object serves a particular function is therefore always a partly observer-

<sup>23</sup> DC 5.1.    <sup>24</sup> EL 27.13; DCv 15.16; 18.4; L 31.38; 570; DC 5.1.

<sup>25</sup> DH 10.2. See Aarsleff (1982: 25).    <sup>26</sup> Searle (1995).    <sup>27</sup> DC 10.7.

relative or artificial and ontologically subjective property: it serves a function only relative to purposes posited by agents. A configuration of wood, for example, functions as a chair relative to purposes posited by humans; its natural properties are insufficient to make it a chair; being a chair depends on there being agents who desire to sit. That a specific object is a chair is nevertheless epistemically objective: that a wooden object could serve agents' desire to sit is independent of their attitudes towards the particular object.

What is distinct about the imposition of a *status* function on a set of objects is that, given a set of agentive purposes constituting the function, the objects' natural properties are neither necessary nor sufficient for them to serve the function. That a particular vocal sound functions as a linguistic sign is a *wholly* artificial property. Semiotic properties are not really the properties of words, but of the intentional agents who take them to be signs.

Propositions about intersubjective meaning, moreover, are truth apt. Indeed, that a given word serves as a sign is epistemically objective for individuals: a speaker may be mistaken about a constitutive intersubjective rule governing the use of a word as sign. In short, on Hobbes's account, linguistic semiotic properties are wholly artificial and therefore ontologically subjective, epistemically objective, actual properties of linguistic signs; and they are real, natural, ontologically subjective, epistemically objective, actual properties of speakers of the language. What is crucial to see here – because significant for Hobbes's account of normative propositions – is that the wholly *artificial*, ontologically subjective properties of a sign can still be the object of truth-apt, *epistemically objective* propositions.

According to Hobbes, signs (and marks) are insufficient for language: to form propositions, language also requires syntactic rules governing the relations between terms.<sup>28</sup> Moreover, there is an important difference between the rules of signification and of syntax. A sound's ability to serve as the sign of a particular conception does not depend on the sound's natural properties. But unlike linguistic signs and their constitutive rules, syntactical rules are not wholly arbitrary and artificial. Some aspects of syntax are artificial – which is why some languages use a word such as 'is' for the copula, while others use word order.<sup>29</sup> But for Hobbes, the syntax of any meaningful language must incorporate rules that are not conventional: rules of syntax include rules of logic. We can approach Hobbes's account of

<sup>28</sup> Abizadeh (2015). <sup>29</sup> L 46.16–17: 1078–1080.

normative properties and propositions by first asking about the epistemic and ontological status of these logical rules.

### Normative Properties

Hobbes was not a conventionalist or “arbitrarist” about logic. The truths articulated by formal logical rules hold independently of anyone’s attitude towards them or the features they concern; they are true independently of anyone’s assent to them or decision to endorse them.<sup>30</sup> Logical rules are therefore natural and are either presupposed or discovered by natural reasoning. Hobbes treated logical rules as both *truth apt* and *epistemically objective*.

One foundational logical rule embedded in syntax echoes the principle of non-contradiction Aquinas had appropriated from Aristotle’s *Metaphysics* 4.9. As Hobbes put it in *De Corpore*, “The certainty of this axiom, viz. of two contradictory names, one is the name of anything whatsoever, the other not, is the original [*principium*] and foundation of all ratiocination, that is, of all philosophy.” The principle’s upshot is that, because “positive names” (such as ‘philosopher’) and “negative names” (such as ‘non-philosopher’) are “contradictory to one another,” they “cannot be the name of the same thing.”<sup>31</sup> Non-contradiction imposes a non-conventional constraint on the artificial creation of conventional rules of signification and syntax. The same is true of the rules governing syllogistic inference, i.e., the rules governing when a “third proposition follows from the two first,” which make it the case that “if those be conceded to be true, it is necessary [*necesse*] that this also be conceded to be true,” or, as Hobbes alternatively put it, “when these being supposed to be true, it cannot be supposed [*supponi non potest*] that the other is not also true.”<sup>32</sup>

Hobbes treated logical rules as not only truth apt, but also *prescriptive*. On the one hand, the “axiom” of non-contradiction articulates a truth known with “certainty.” On the other hand, non-contradiction consists in a *rule* prescribing to agents not to use a positive name and the corresponding negative name for the same object. To generalize: logical rules prescribe to agents how to reason in language and what sets of propositions to assent to or suppose.

The question is how a *prescription* could be true or false. The answer is that Hobbes did not take logical rules merely to *prescribe* which

<sup>30</sup> Malcolm (2002: 152); Jesseph (1999: chapter 5). <sup>31</sup> DC 2.8; cf. ST I-II 94.2.

<sup>32</sup> DC 4.1; 3.18.

propositions to assent to or how to reason; he took the prescriptions to be genuinely *normative*. This is part of the force of *necesse* and *non potest* in the passages previously cited. In virtue of the principle of non-contradiction, an agent has a normative reason not to assent to or suppose both the proposition 'Socrates is a philosopher' and the proposition 'Socrates is a non-philosopher'. Hobbes took purported rules of logic to be truth apt because they concern reason relations: they articulate, for example, the truth that agents have an epistemic reason to avoid assenting to contradictory propositions.

Reason relations, in turn, concern the normative properties of objects. But these normative properties do not have causal standing: that a fact provides an agent with a normative reason to favour something, does not imply that the fact itself will actually motivate the agent to favour it. And if normative properties have no causal standing, then they are not *real* properties. That something is pleasant for agents and is therefore disposed to cause them to *see* it as good is a real property of the object; that it is taken to be good is also a real property of the object, even though it is really the property of the takers; but that it *is* good for agents is not a real property of anything. On Hobbes's account, reason relations and normative properties are irreal.

Normative propositions can nevertheless still be the object of truth-apt and even epistemically objective propositions. Just as it is true that Don Quixote is mad, and just as the proposition that he is mad can be epistemically objective, so too can it be true and epistemically objective that one has a reason to secure self-preservation. Nor did the irreal character of normative properties prevent Hobbes from taking some to be *natural* properties that obtain independently of being represented to obtain. Individuals in the state of nature would still have a prudential reason to be disposed to show gratitude, or to avoid abusing alcohol, even if all of them were Fooles or Cyrenaics who failed to conceive of peaceful or temperate dispositions as normative. This is why natural laws are *natural*: they are normative in virtue of properties obtaining independently of agential representation. Similarly, some things are good for agents even though they fail to recognize their goodness – they are good by nature.

A crucial difference between normative properties constituted by reasons of the good and those constituted by reasons of the right is that whereas the former can be natural, on Hobbes's account *all normative properties of the right are artificial*. This is because of two features of Hobbes's ethics. First, in order to have reasons of the right, one must be

a person, and to be a person, one must be represented as a person. Second, in order for persons to have such reasons, they must *acquire* them, and to do so they must be represented as having willingly acquired them.

Even irreal *artificial* normative properties can be the subject of truth-apt, epistemically objective propositions. Just as the proposition that ‘rain’ signifies a conception of rain is (on Hobbes’s account) true and epistemically objective (even though being a linguistic sign is an artificial property of ‘rain’), so too can the proposition that subjects are obligated to obey their sovereign. Yet this normative fact depends on the subjects being represented as persons and being represented as having covenanted to obey. (If no one, including the sovereign, represents a subject as obligated anymore, the obligation lapses.) It is an epistemically objective, true fact that subjects are obligated to obey, but this truth concerns irreal, artificial normative properties.

There is, however, a crucial difference between non-normative artificial properties, whether real or irreal, and irreal normative properties, whether natural or artificial – a difference bearing on the epistemological question of how the truth of normative propositions could be known. That ‘rain’ is a sign of a conception of rain is the word’s real, artificial property, which can be reductively analyzed in terms of the real, natural properties of agents. These properties therefore have causal standing. By contrast, being good or obligatory is not reducible to the real property of anything: there is no causal contact. The implication is that no one can know normative truths solely via causal contact with the world, such as via sensory perception. Don Quixote’s madness is also an irreal (because purely fictional) property, but we can know of his madness because it is the irreal property of an irreal object *created* by intentional agents: because he is mad just in virtue of being represented as mad, to know that he is mad it is sufficient to know how he is represented by the relevant agents – and how Don Quixote is represented is a real property with causal standing. But foundational normative precepts such as the Instrumental-Transmission, Prudential, and Juridical Precepts – like the principle of non-contradiction – are not true in virtue of being represented as true.

That in Hobbes’s view foundational normative truths cannot be known solely via sensory perception or other causal contact with the world can be discerned from the contrast he drew between humans and beasts. Although nonhuman animals have desires, and see the objects of their desires *as good*, and even though some things may satisfy their desires, nothing is properly *normative* for them. Nonhumans cannot *have* reasons because reasons, for Hobbes, must be epistemically accessible – which they are not for creatures

incapable of linguistic reasoning. This is why Hobbes's "empiricism" – his claim that sensory perception is the ultimate source of human knowledge<sup>33</sup> – does not gainsay the claim that normative truths are known via something other than causal contact: we discover reasons in part via our awareness of the constitutive conditions of rational agency and of reasoning. This is ultimately why Hobbes could construe ethics as a science of the mind: ethics concerns the properties, including the normative properties, of our linguistically mediated mental life *qua* rational agents.

The epistemic standard Hobbes invoked here is, as I suggested in [Chapter 3](#), *conceivability* or intelligibility. This is the standard in terms of which Hobbes conceived and justified deterministic materialism: he defined the notion of a cause in terms of it.<sup>34</sup> Although Hobbes portrayed scientific reasoning as about causes, he took it for granted that science must ultimately be grounded in foundational propositions that are self-evidently true in virtue of the fact that, on reflection, rational agents could not but conceive them to be true. The fundamental ontological truths of Hobbes's corpuscularian materialism – according to which all effects are caused by bodies in motion – "are manifest of themselves [*manifestae sunt per se*], or (as they say commonly) known to nature [*naturae (ut dicunt) nota*]; so that they need no method [i.e., analysis] at all."<sup>35</sup> Such "universal principles . . . cannot be demonstrated; and seeing they are known by nature [*sint nota naturae*] . . . need no demonstration, though they need explication."<sup>36</sup> We do not discover these fundamental ontological truths via syllogistic reasoning; they are "manifest" to us in virtue of our inability to conceive of them as not true.

Hobbes deployed the conceivability standard not only for knowledge of the mechanical causes of our sensory perceptions, but also for our implicit knowledge of the pre-conventional logical norms embedded in linguistic syntax. For example, he glossed his initial formulation of logical inference quoted above ("A proposition is said to *follow* from two other propositions, when these being supposed to be true, it cannot be supposed that the other is not also true") as equivalent to saying that a conclusion is said "to be necessarily *inferred* from" two propositions when "it is not intelligible [*intelligi non potest*]" that the conclusion not also be true if the former are taken to be true.<sup>37</sup> Conceivability is the epistemic standard by which the *normative* "necessity" of the logical inference is also known. Similarly, I may be able to discover that peace is a relevant means to self-preservation via reasoning that draws on causal information, but my knowledge of the

<sup>33</sup> L 1.2. 22.    <sup>34</sup> DC 6.10.    <sup>35</sup> DC 6.5.    <sup>36</sup> DC 6.12.    <sup>37</sup> DC 3.18.

foundationally normative Prudential Precepts, which undergird the normativity of the laws of nature, depends on the insight that I cannot conceive of myself as a rational agent without also conceiving that I reason in order to care for my ongoing good. In conceiving myself as a scientific reasoner, I come to know that “*Reason is the pace; Encrease of Science, the way; and the Benefit of man-kind, the end.*”<sup>38</sup> Similarly, my knowledge of the foundationally normative Juridical Precepts depends on the insight that I cannot conceive of myself as a person unless I take myself as able to acquire directed obligations in producing reliable signs of an intention to do so.

Just as rational agents cannot conceive of real objects in the world without extension and motion, they cannot conceive of beliefs, desires, and reasoning, nor actually believe, desire, or engage in reasoning, without reasons. Normativity is the proper subject of scientific inquiry for Hobbes because it concerns properties rational agents cannot conceivably do without. Much of this set of metaethical commitments is implicit rather than fully worked out in Hobbes. But this implicit framework gives the lie to efforts to find in Hobbes’s ethical naturalism a reductionist or noncognitivist metaethics.

Hobbes’s significance in the history of ethics is therefore twofold. At the level of metaethics, he was at the forefront of attempts to reconcile the new materialist, mechanistic science with irreducibly normative truths and properties. At the level of normative ethics, the new notion of obligation articulated by Suárez, Grotius, and Hobbes lay the foundations for the modern, juridical conception of ethics that flourished in the subsequent century; Hobbes inaugurated the voluntarist strand that grounds obligation in individuals’ own, mutually intelligible actions. And Hobbes’s model of juridical obligation – unlike subsequent incarnations that wholly sever it from eudaimonistic concerns – remained tightly linked to and shaped by prudential concerns about one’s own ongoing good.

<sup>38</sup> L 5.18: 74.

## Works Cited

All underlining in quotations are my additions, *italics* in original. Translations from Latin are frequently modified. For primary texts, see Abbreviations list. References to primary texts follow the format: [volume.]chapter. paragraph: pages.

- Aarsleff, Hans. 1982. *From Locke to Saussure*. Minneapolis: University of Minnesota Press.
- Abizadeh, Arash. 2011. "Hobbes on the Causes of War." *American Political Science Review* 105.2:298–315.
2013. "Leviathan as Mythology." In *Hobbes Today*, ed. S. A. Lloyd. Cambridge: Cambridge University Press.
2015. "The Absence of Reference in Hobbes's Philosophy of Language." *Philosophers' Imprint* 15.22:1–17.
2016. "Sovereign Jurisdiction, Territorial Rights, and Membership in Hobbes." In *The Oxford Handbook of Hobbes*, ed. A. P. Martinich and K. Hoekstra. New York: Oxford University Press.
- 2017a. "Hobbes's Conventionalist Theology, the Trinity, and God as an Artificial Person by Fiction." *The Historical Journal* 60.4:915–941.
- 2017b. "Hobbes on Mind." *Journal of the History of Philosophy* 55.1:1–34.
- 2017c. "Hobbes's Agnostic Theology before *Leviathan*." *Canadian Journal of Philosophy* 47.5:714–737.
- In progress. "Subjectivism, Instrumentalism, and Prudentialism about Reasons."
- Annas, Julia. 1993. *The Morality of Happiness*. Oxford: Oxford University Press.
- Anscombe, G. E. M. 1958. "Modern Moral Philosophy." *Philosophy* 33.124:1–19.
- Barry, Brian. 1972. "Warrender and His Critics." In *Hobbes and Rousseau*, ed. M. Cranston and R. S. Peters. Garden City: Anchor Books.
- Baumgold, Deborah. 1983. "Subjects and Soldiers." *History of Political Thought* 4.1:43–64.
- Blau, Adrian. 2016. "Reason, Deliberation, and the Passions." In *The Oxford Handbook of Hobbes*, ed. A. P. Martinich and K. Hoekstra. Oxford: Oxford University Press.
2017. "Interpreting Texts." In *Methods in Analytical Political Theory*. Cambridge: Cambridge University Press.
- Bobbio, Norberto. 1993. *Thomas Hobbes and the Natural Law Tradition*. Chicago: University of Chicago Press.

- Boonin-Vail, David. 1994. *Thomas Hobbes and the Science of Moral Virtue*. Cambridge: Cambridge University Press.
- Brandt, Frithiof. 1927. *Thomas Hobbes' Mechanical Conception of Nature*. Copenhagen: Levin & Munksgaard.
- Brandt, Richard. 1979. *A Theory of the Good and the Right*. Oxford: Clarendon Press.
1992. *Morality, Utilitarianism, and Rights*. Cambridge: Cambridge University Press.
- Broome, John. 1994. "Discounting the Future." *Philosophy & Public Affairs* 23.2:128–156.
2005. "Does Rationality Give us Reasons?" *Philosophical Issues* 15.1:321–337.
- 2007a. "Is Rationality Normative?" *Disputatio* 2.23:161–178.
- 2007b. "Wide or Narrow Scope?" *Mind* 116.462:359–370.
2013. *Rationality Through Reasoning*. Oxford: Wiley Blackwell.
- Burnyeat, M. F. 1992. "Is an Aristotelian Philosophy of Mind Still Credible?" In *Essays on Aristotle's De Anima*, ed. M. Nussbaum and A. O. Rorty. Oxford: Oxford University Press.
- Caston, Victor. 2005. "The Spirit and the Letter." In *Metaphysics, Soul, and Ethics*, ed. R. Salles. Oxford: Clarendon Press.
- Cohen, G. A. 1996. "Reason, Humanity, and the Moral Law." In *The Sources of Normativity*, ed. O. O'Neill. Cambridge: Cambridge University Press.
- Copp, David. 1980. "Hobbes on Artificial Persons and Collective Actions." *The Philosophical Review* 89.4:579–606.
- Crane, Tim. 2001. *Elements of Mind*. Oxford: Oxford University Press.
- Crisp, Roger. 2006. *Reasons and the Good*. Oxford: Oxford University Press.
- Cromartie, Alan. 2011. "The Elements and Hobbesian Moral Thinking." *History of Political Thought* 32.1:21–47.
- Cuffaro, Michael. 2011. "On Thomas Hobbes's Fallible Natural Law Theory." *History of Philosophy Quarterly* 28.2:175–190.
- Curley, Edwin. 2004. "The Covenant with God in Hobbes's *Leviathan*." In *Leviathan After 350 Years*, ed. T. Sorell and L. Foisneau. Oxford: Clarendon Press.
- Darwall, Stephen. 1983. *Impartial Reason*. Ithaca: Cornell University Press.
1992. "Internalism and Agency." *Philosophical Perspectives* 6:155–174.
1995. *The British Moralists and the Internal 'Ought'*. Cambridge: Cambridge University Press.
1998. *Philosophical Ethics*. Boulder: Westview Press.
2000. "Normativity and Projection in Hobbes's *Leviathan*." *The Philosophical Review* 109.3:313–347.
2009. *The Second-Person Standpoint*. Cambridge: Harvard University Press.
2012. "Grotius at the Creation of Modern Moral Philosophy." *Archiv für Geschichte der Philosophie* 94.3:296–325.
- Davidson, Donald. 1984. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
- de Jong, Willem. 1986. "Hobbes's Logic." *History and Philosophy of Logic* 7.2:123–142.

- Deigh, John. 1996. "Reason and Ethics in Hobbes's *Leviathan*." *Journal of the History of Philosophy* 34.1:33–60.
- Duff, R. A. 2001. *Punishment, Communication, and Community*. Oxford: Oxford University Press.
- Dyzenhaus, David. 2001. "Hobbes and the Legitimacy of Law." *Law and Philosophy* 20.5:461–498.
2010. "Hobbes's Constitutional Theory." In *Leviathan*, ed. I. Shapiro. New Haven: Yale University Press.
2012. "Hobbes on the Authority of Law." In *Hobbes and the Law*, ed. D. Dyzenhaus and T. Poole. Cambridge: Cambridge University Press.
- Engel, Srul. 1959. *The Philosophy of Language in Hobbes and Locke*. PhD Thesis, University of Toronto, Toronto.
1962. "Analogy and Equivocation in Hobbes." *Philosophy* 37.142:326–335.
- Feinberg, Joel. 1965. "The Expressive Function of Punishment." *The Monist* 49.3:397–423.
- Forsberg, Ralph. 1990. *Thomas Hobbes' Theory of Obligation*. Wakefield: Longwood Academic.
- Fox-Decent, Evan. 2012. "Hobbes's Relational Theory." In *Hobbes and the Law*, ed. D. Dyzenhaus and T. Poole. Cambridge: Cambridge University Press.
- Friedle, Simon. 2012. *Thomas Hobbes and the Reception of Early-Modern Epicureanism*. PhD Thesis, Faculty of History, University of Cambridge.
- Frost, Samantha. 2008. *Lessons from a Materialist Thinker*. Stanford: Stanford University Press.
- Fuller, Lon. 1969. *The Morality of Law*. Revised ed. New Haven: Yale University Press.
- Garrett, Aaron. 2013. "Seventeenth-Century Moral Philosophy." In *The Oxford Handbook of the History of Ethics*, ed. R. Crisp. Oxford: Oxford University Press.
- Garsten, Bryan. 2010. "Religion and Representation in Hobbes." In *Leviathan*, ed. I. Shapiro. New Haven: Yale University Press.
- Gauthier, David. 1969. *The Logic of Leviathan*. Oxford: Clarendon Press.
1979. "Thomas Hobbes: Moral Theorist." *The Journal of Philosophy* 76.10:547–559.
- Geach, Peter. 1962. *Reference and Generality*. Ithica: Cornell University Press.
- Gert, Bernard. 1965. "Hobbes, Mechanism, and Egoism." *The Philosophical Quarterly* 15.61:341–349.
1967. "Hobbes and Psychological Egoism." *Journal of the History of Ideas* 28.4:503–520.
2001. "Hobbes on Reason." *Pacific Philosophical Quarterly* 82.3–4:243–257.
- Gibson, Martha. 2004. *From Naming to Saying*. Oxford: Blackwell.
- Gordlberg, John, and Benjamin Zipursky. 2010. "Torts as Wrongs." *Texas Law Review* 88.5:917–986.
- Green, Michael. 2015. "Authorization and Political Authority in Hobbes." *Journal of the History of Philosophy* 53.1:25–47.
- Grice, H. P. 1968. "Utterer's Meaning, Sentence-Meaning, and Word-Meaning." *Foundations of Language* 4.3:225–242.

- Hacking, Ian. 2006. *The Emergence of Probability*. 2nd ed. Cambridge: Cambridge University Press.
- Hampton, Jean. 1986. *Hobbes and the Social Contract Tradition*. Cambridge: Cambridge University Press.
1992. "Hobbes and Ethical Naturalism." *Philosophical Perspectives* 6:333–353.
1998. *The Authority of Reason*. Cambridge: Cambridge University Press.
- Hare, R. M. 1952. *The Language of Morals*. Oxford: Clarendon Press.
- Harman, Gilbert. 1977. *The Nature of Morality*. Oxford: Oxford University Press.
- Hart, H. L. A. 1961. *The Concept of Law*. Oxford: Clarendon Press.
2008. *Punishment and Responsibility*. 2nd ed. Oxford: Oxford University Press.
- Heathwood, Chris. 2005. "The Problem of Defective Desires." *Australasian Journal of Philosophy* 83.4:487–504.
- Hoekstra, Kinch. 1997. "Hobbes and the Foole." *Political Theory* 25.5:620–654.
2003. "Hobbes on Law, Nature, and Reason." *Journal of the History of Philosophy* 41.1:111–120.
- Hohfeld, Wesley Newcomb. 1964. *Fundamental Legal Conceptions, as Applied in Judicial Reasoning*. Ed. W. W. Cook. New ed. New Haven: Yale University Press.
- Holden, Thomas. 2004. *The Architecture of Matter*. Oxford: Oxford University Press.
2015. "Hobbes's First Cause." *Journal of the History of Philosophy* 53.4:647–667.
2016. "Hobbes on the Function of Evaluative Speech." *Canadian Journal of Philosophy* 46.1:123–144.
- Hubin, Donald C. 1999. "What's Special about Humeanism." *Noûs* 33.1:30–45.
- Hungerland, Isabel, and George Vick. 1981. "Hobbes's Theory of Language, Speech, and Reasoning." In *Computatio Sive Logica*, ed. I. Hungerland and G. Vick. Norwalk: Abaris Books.
- Irwin, Terence. 2008. *The Development of Ethics: Volume 2*. Oxford: Oxford University Press.
- Jackson, Nicholas D. 2007. *Hobbes, Bramhall and the Politics of Liberty and Necessity*. Cambridge: Cambridge University Press.
- Jesseph, Douglas. 1996. "Hobbes and the Method of Natural Science." In *The Cambridge Companion to Hobbes*, ed. T. Sorell. Cambridge: Cambridge University Press.
1999. *Squaring the Circle*. Chicago: University of Chicago Press.
2010. "Scientia in Hobbes." In *Scientia in Early Modern Philosophy*, ed. T. Sorell, G. A. Rogers and J. Krayer. Dordrecht: Springer.
- Johnston, David. 1986. *The Rhetoric of Leviathan*. Princeton: Princeton University Press.
- Kahn, Victoria. 1985. *Rhetoric, Prudence, and Skepticism in the Renaissance*. Ithaca: Cornell University Press.
- Kavka, Gregory. 1986. *Hobbesian Moral and Political Theory*. Princeton: Princeton University Press.
- Kemp, J. 1970. *Ethical Naturalism*. London: MacMillan.
- Kiesewetter, Benjamin. 2015. "Instrumental Normativity." *Ethics* 125.4:921–946.

- Kolodny, Niko. 2005. "Why Be Rational?" *Mind* 114.455:509–563.
- Korsgaard, Christine M. 1996a. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- 1996b. *The Sources of Normativity*. Ed. O. O'Neill. Cambridge: Cambridge University Press.
2007. "Autonomy and the Second Person Within." *Ethics* 118.1:8–23.
2008. "The Normativity of Instrumental Reason." In *The Constitution of Agency*. Oxford: Oxford University Press.
- Lagerlund, Henrik. 2008. "Mental Representation in Medieval Philosophy." In *The Stanford Encyclopedia of Philosophy*. Ed. E. N. Zalta. Fall 2008 ed.
- LeBar, Mark. 2009. "Virtue Ethics and Deontic Constraints." *Ethics* 119.4:642–671.
- Leijenhorst, Cees. 2002. *The Mechanisation of Aristotelianism*. Leiden: Brill.
- Lin, Eden. 2018. "Asymmetrism about Desire Satisfaction and Time." In *Oxford Studies in Normative Ethics Volume 7*, ed. M. Timmons. Oxford: Oxford University Press.
- Lloyd, S. A. 1992. *Ideals as Interests in Hobbes's Leviathan*. Cambridge: Cambridge University Press.
2009. *Morality in the Philosophy of Thomas Hobbes*. Cambridge: Cambridge University Press.
- Losonsky, Michael. 2001. *Enlightenment and Action from Descartes to Kant*. Cambridge: Cambridge University Press.
- Lott, Tommy. 1982. "Hobbes's Mechanistic Psychology." In *Thomas Hobbes*, ed. J. G. van der Bend. Amsterdam: Editions Rodopi B. V.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin Books.
- Malcolm, Noel. 2002. *Aspects of Hobbes*. Oxford: Clarendon Press.
- Martinich, A. P. 1992. *The Two Gods of Leviathan*. Cambridge: Cambridge University Press.
2005. *Hobbes*. New York: Routledge.
2016. "Authorization and Representation in Hobbes's *Leviathan*." In *The Oxford Handbook of Hobbes*, ed. A. P. Martinich and K. Hoekstra. New York: Oxford University Press.
- Matheron, Alexandre. 1990. "Hobbes, la Trinité et les caprices de la représentation." In *Thomas Hobbes*, ed. Y. C. Zarka and J. Bernhardt. Paris: Presses Universitaires de France.
- May, Simon Căbulea. 2012. "Moral Status and the Direction of Duties." *Ethics* 123.1:113–128.
- McNeilly, F.S. 1968. *The Anatomy of Leviathan*. New York: St. Martin's Press.
- Metzger, Ernest. 1997. "Actions." In *A Companion to Justinian's Institutes*. New York: Cornell University Press.
- Meyers, C. D. 2013. "Hobbes and the Rationality of Self-Preservation." *The European Legacy* 18.3:269–286.
- Miller, Alexander. 2003. *An Introduction to Contemporary Metaethics*. Cambridge: Polity Press.

- Mintz, Samuel. 1969. *The Hunting of Leviathan*. Cambridge: Cambridge University Press.
- Mitsis, Phillip. 1988. *Epicurus' Ethical Theory*. Ithaca: Cornell University Press.
- Moore, G. E. 1922. "The Nature of Moral Philosophy." In *Philosophical Studies*. London: Routledge & Kegan Paul.
- Murphy, Mark. 1994. "Deviant Uses of 'Obligation' in Hobbes' *Leviathan*." *History of Philosophy Quarterly* 11.3:281–294.
1995. "Was Hobbes a Legal Positivist?" *Ethics* 105.4:846–873.
2000. "Hobbes on the Evil of Death." *Archiv für Geschichte der Philosophie* 82.1:36–61.
- Nagel, Thomas. 1959. "Hobbes's Concept of Obligation." *The Philosophical Review* 68.1:68–83.
1970. *The Possibility of Altruism*. Princeton: Princeton University Press.
- Nuchelmans, Gabriel. 1983. *Judgment and Proposition*. Amsterdam: North-Holland Publishing.
- Oakeshott, Michael. 1946. "Introduction." In *Leviathan*. Oxford: Basil Blackwell.
- Olsthoorn, Johan. 2014. "Worse than Death." *Hobbes Studies* 27.2:148–170.
2015. "Why Justice and Injustice Have No Place Outside the Hobbesian State." *European Journal of Political Theory* 14.1:19–36.
- Ott, Walter. 2004. *Locke's Philosophy of Language*. Cambridge: Cambridge University Press.
- Paganini, Gianni. 2012. "'Passionate Thought': Reason and the Passion of Curiosity in Thomas Hobbes." In *Emotional Minds*, ed. S. Ebbsermeyer. Berlin: De Gruyter.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
2011. *On What Matters*. 2 vols. Oxford: Oxford University Press.
- Parkin, Jon. 2007. *Taming the Leviathan*. Cambridge: Cambridge University Press.
- Patin, Gui. 1846. *Lettres de Gui Patin: Nouvelle édition augmentée de lettres inédites*. Ed. J.-H. Reveillé-Parise. Vol. 2. Paris: J.-B. Ballière.
- Pécharman, Martine. 1992. "Le Discours mental selon Hobbes." *Archives de Philosophie* 55.4:553–573.
- Peters, Richard. 1956. *Hobbes*. London: Penguin.
- Pettit, Philip. 2008. *Made with Words*. Princeton: Princeton University Press.
- Pink, Thomas. 2004. "Suarez, Hobbes and the Scholastic Tradition in Action Theory." In *The Will and Human Action*, ed. T. Pink and M. W. F. Stone. London: Routledge.
2010. "Thomas Hobbes." In *A Companion to the Philosophy of Action*, ed. T. O'Connor and C. Sandis. Oxford: Blackwell.
2011. "Thomas Hobbes and the Ethics of Freedom." *Inquiry* 54.5:541–563.
- Pitkin, Hanna. 1967. *The Concept of Representation*. Berkeley: University of California Press.
- Plamenatz, John. 1965. "Mr. Warrender's Hobbes." In *Hobbes Studies*, ed. K. C. Brown. Oxford: Basil Blackwell.

- Pocock, J. G. A. 1997. "Time, History and Eschatology in the Thought of Thomas Hobbes." In *Hobbes*, ed. J. Dunn and I. Harris. Cheltenham: Edwin Elgar Publishing Limited.
- Railton, Peter. 1989. "Naturalism and Prescriptivity." *Social Philosophy and Policy* 7.01:151–174.
1997. "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action." In *Ethics and Practical Reason*, ed. G. Cullity and B. Gaut. Oxford: Clarendon Press.
- Raphael, D. D. 1977. *Hobbes: Morals and Politics*. London: George Allen & Unwin Ltd.
- Raz, Joseph. 2005. "The Myth of Instrumental Rationality." *Journal of Ethics & Social Philosophy* 1.1:2–28.
- Riley, Patrick. 1982. *Will and Political Legitimacy*. Cambridge: Harvard University Press.
- Ripstein, Arthur. 1987. "Foundationalism in Political Theory." *Philosophy & Public Affairs* 16.2:115–137.
- Ross, George MacDonald. 1987. "Hobbes's Two Theories of Meaning." In *The Figural and the Literal*, ed. A. Benjamin, G. Cantor, and J. Christie. Manchester: Manchester University Press.
- Runciman, David. 2000. "What Kind of Person Is Hobbes's State?" *The Journal of Political Philosophy* 8.2:268–278.
2009. "Hobbes's Theory of Representation." In *Political Representation*, ed. I. Shapiro, S. Stokes, E. J. Wood, and A. Kirshner. Cambridge: Cambridge University Press.
- Rutherford, Donald. 2003. "In Pursuit of Happiness." *Philosophical Topics* 31.1/2:369–393.
2012. "The End of Ends?" In *The Reception of Aristotle's Ethics*, ed. J. Miller. Cambridge: Cambridge University Press.
- Sarch, Alexander. 2013. "Desire Satisfactionism and Time." *Utilitas* 25.2:221–245.
- Scanlon, Thomas. 1998. *What We Owe Each Other*. Cambridge: Harvard University Press.
2014. *Being Realistic About Reasons*. Oxford: Oxford University Press.
- Schaffer, Simon. 1988. "Wallfaction." *Studies in History and Philosophy of Science* 19.3:275–298.
- Schneewind, J. B. 1998. *The Invention of Autonomy*. Cambridge: Cambridge University Press.
- Schroeder, Mark. 2007. *Slaves of the Passions*. Oxford: Oxford University Press.
- Schuhmann, Karl. 2004. "Phantasms and Idols." *Rivista di storia della filosofia* 59.1:15–31.
- Searle, John. 1995. *The Construction of Social Reality*. New York: The Free Press.
- Shoemaker, David. 2011. "Attributability, Answerability, and Accountability." *Ethics* 121.3:602–632.
- Sidgwick, Henry. 1896. *Outlines of the History of Ethics for English Readers*. 4th ed. London: Macmillan.

- Simmons, A. John. 1979. *Moral Principles and Political Obligation*. Princeton: Princeton University Press.
- Skinner, Quentin. 1988. *Meaning and Context*. Ed. J. Tully. Princeton: Princeton University Press.
1996. *Reason and Rhetoric in the Philosophy of Hobbes*. Cambridge: Cambridge University Press.
1999. "Hobbes and the Purely Artificial Person of the State." *The Journal of Political Philosophy* 7.1:1–29.
2002. *Visions of Politics Volume III*. Cambridge: Cambridge University Press.
2005. "Hobbes on Representation." *European Journal of Philosophy* 13.2:155–184.
- Skorupski, John. 2007. "Internal Reasons and the Scope of Blame." In *Bernard Williams*, ed. A. Thomas. Cambridge: Cambridge University Press.
2010. *The Domain of Reasons*. Oxford: Oxford University Press.
- Slopp, Gabriella. 1998. "From Genus to Species." *History of Political Thought* 19.4:552–569.
2007. "Hobbes on Glory and Civil Strife." In *The Cambridge Companion to Hobbes's Leviathan*, ed. P. Springborg. Cambridge: Cambridge University Press.
- Smart, J. J. C. 1961. "Free-Will, Praise and Blame." *Mind* 70.279:291–306.
- Smith, Angela. 2005. "Responsibility for Attitudes." *Ethics* 115.2:236–271.
2007. "On Being Responsible and Holding Responsible." *The Journal of Ethics* 11.4:465–484.
- Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.
- Sobel, David. 2001. "Subjective Accounts of Reasons for Action." *Ethics* 111.3:461–492.
- Sorabji, Richard. 1974. "Body and Soul in Aristotle." *Philosophy* 49.187:63–89.
- Sorell, Tom. 1986. *Hobbes*. London: Routledge & Kegan Paul.
- Sreedhar, Susanne. 2010. *Hobbes on Resistance*. Cambridge: Cambridge University Press.
- Sreenivasan, Gopal. 2010. "Duties and Their Direction." *Ethics* 120.3:465–494.
- Stoffell, Brian. 1991. "Hobbes on Self-Preservation and Suicide." *Hobbes Studies* 4.1:26–33.
- Strauss, Leo. 1963. *The Political Philosophy of Hobbes*. Chicago: University of Chicago Press.
- Strawson, P. F. 2008. *Freedom and Resentment and Other Essays*. London: Routledge.
- Taylor, A. E. 1965. "The Ethical Doctrine of Hobbes." In *Hobbes Studies*, ed. K. C. Brown. Oxford: Basil Blackwell.
- Thompson, Michael. 2004. "What Is It to Wrong Someone?" In *Reason and Value*, ed. R. J. Wallace, P. Pettit, S. Scheffler, and M. Smith. Oxford: Clarendon Press.
- Tralau, Johan. 2011. "Hobbes contra Liberty of Conscience." *Political Theory* 39.1:58–84.
- Tuck, Richard. 1979. *Natural Rights Theories*. Cambridge: Cambridge University Press.

1987. "The 'Modern' Theory of Natural Law." In *The Languages of Political Theory in Early-Modern Europe*, ed. A. Pagden. Cambridge: Cambridge University Press.
- 1988a. "Hobbes and Descartes." In *Perspectives on Thomas Hobbes*, ed. G. A. J. Rogers and A. Ryan. Oxford: Clarendon Press.
- 1988b. "Optics and Sceptics." In *Conscience and Casuistry in Early Modern Europe*, ed. E. Leites. Cambridge: Cambridge University Press.
1989. *Hobbes*. Oxford: Oxford University Press.
1996. "Hobbes's Moral Philosophy." In *The Cambridge Companion to Hobbes*, ed. T. Sorell. Cambridge: Cambridge University Press.
- Tweedale, Martin. 1990. "Mental Representation in Later Medieval Scholasticism." In *Historical Foundations of Cognitive Science*, ed. J.-C. Smith. Dordrecht: Kluwer Academic Publishing.
- van Apeldoorn, Laurens. 2012. "Reconsidering Hobbes's Account of Practical Deliberation." *Hobbes Studies* 25.2:143–165.
- Vieira, Mónica Brito. 2009. *The Elements of Representation in Hobbes*. Leiden: Brill.
- Wallace, R. Jay. 1996. *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.
- Warrender, Howard. 1957. *The Political Philosophy of Hobbes*. Oxford: Clarendon Press.
- Watkins, J. W. N. 1955. "Philosophy and Politics in Hobbes." *The Philosophical Quarterly* 5.19:125–146.
1973. *Hobbes's System of Ideas*. 2nd ed. London: Hutchinson University Library.
- Watson, Gary. 1996. "Two Faces of Responsibility." *Philosophical Topics* 24.2:227–248.
2004. *Agency and Answerability*. Oxford: Oxford University Press.
- Williams, Bernard. 1981. *Moral Luck*. Cambridge: Cambridge University Press.
- Wolfsdorf, David. 2013. *Pleasure in Ancient Greek Philosophy*. Cambridge: Cambridge University Press.
- Wolin, Sheldon S. 2006. *Politics and Vision*. Expanded ed. Princeton: Princeton University Press.
- Wright, Crispin. 1992. *Truth and Objectivity*. Cambridge: Harvard University Press.
- Yates, Arthur. 2013. "A Hohfeldian Analysis of Hobbesian Rights." *Law and Philosophy* 32.4:405–434.
- Zarka, Yves Charles. 1996. "First Philosophy and the Foundations of Knowledge." In *The Cambridge Companion to Hobbes*, ed. T. Sorell. Cambridge: Cambridge University Press.
1999. *La Décision métaphysique de Hobbes*. 2nd expanded ed. Paris: Vrin.

# Index

- accountability, 2–6, 9–11, 21–23, 181, 187–191, 205, 209–219, 223, 226–227, 236–237, 239–241, 244, 246, 258–262
- holding accountable, 21, 199–200, 202–203, 208–212
- legal, 21, 208–212, 239, 241, *See also actio*
- standing to hold accountable, 4–6, 11, 21, 190, 198, 199–200, 211–213, 217, 219, 246, 259
- actio* (law of actions), 4, 241
- action
- imputed / attributed, 183, 186, 249, *See also* author
  - voluntary, 18, 31, 62–72, 155, 183, 184, 212–214, 248–255, 258, *See also* blame; responsibility; egoism, psychological; hedonism, psychological; reasons, motivational
- agency, 62–65
- rational, 5, 9, 19, 23, 63, 186–187, 246, 248–249, 252–255, 257
  - voluntary / intentional, 184–187, 248–249, 257
- Aquinas, Thomas, 1–5, 97–99, 110, 118, 183, 206, 272
- Aristotle / Aristotelianism, 1, 27, 36, 81, 97, 128, 139, 142, 147–148, 168, 183–184, 202, 206, 272
- associationalism, mental, 64, 72, 74, 86–90
- attitudes
- propositional, 73–74, 76, 86
  - reactive, 200–205, 207–208, 210, *See also* blame, reactive; passions, reactive
- author / owner, 23, 186, 249–260
- authorization, 23, 41, 196, 197, 209, 216, 238–241, 245–246, 253–260
- blame, 4, 183–191, 212, 219
- critical, 6, 181, 186–187, 188, 201, 203, 252
  - nonnormative, 184, 185, 201, 203, 208, 248
  - normative, 19, 21, 127
  - reactive, 6, 21, 181, 188–189, 190, 200–206, 207–208, 217, 228
- Boonin, David, 106, 227
- Bramhall, John, 67–69, 183–187
- Broome, John, 50
- Carneades, 1–3, 7, 142
- Cicero, 1, 3, 139, 142, 146, 149, 157, 167–168
- civil law, 22, 178, 189, 215, 218, 226, 235–241
- cognitivism, 39, 52
- Cohen, G. A., 260–262
- command, 1–2, 5, 10, 55–57, 91, 178–179, 189–191, 194, 217–218, 231–232, 235–240
- conceivability, 8, 118–119, 264, 275–276
- conscience, 119–124, 188, 212–216
- consent, sign theory of, 7, 22, 219, 233
- counsel, 2, 4, 5, 18–19, 45, 55–57, 60, 91–93, 165, 178, 181, 185–191, 203, 246
- cruelty, 43, 103, 123, 205–209, 237
- Cyrenaics, 109, 117, 139–141, 146, 149, 156–157, 164, 273
- Darwall, Stephen, 30, 32, 36, 46, 117, 190
- definitions, reforming (scientific), 17–18, 37–38, 39, 142–143, 168, 266, *See also* good, scientific meaning of
- Deigh, John, 113–116
- deliberation, 18, 53, 62–72, 76, 103–105, 130, 183, 248–249
- descriptivism, 39, 50, 52–59
- desire (appetite), 31, 33–34, 40–43, 47, 55–56, 57–58, 65–67, 69–70, 87, 145–147, *See also* subjectivism; passions; pleasure of anticipation; reasons for desire
- egoism
- ethical, 1, 110
  - psychological, 117, 146–147, 229
- Epicurus / Epicureans, 3, 20, 139–150, 155–159, 218
- error theory, 8, 12, 17–18, 28–39
- eudaimonia, 1, 2, 139, *See also* felicity

- eudaimonism, 1–4, 6, 110, 139–141, 156–157, 164, 219–220, *See also* obligations, eudaimonistic
- felicity, 2–3, 19–21, 43, 98–100, 109–113, 117–118, 132, 135–141, 149, 156–172, 187, 188, 193, 227, *See also* eudaimonia
- Foole, the, 6, 17, 49, 92–93, 125–130, 137, 166, 167, 177, 224–225, 273
- free will, 62, 65, 67, 183–184, 186, 213
- glory, 43, 100, 160–161, 163, 206, 267–268
- God, 1–2, 4, 5, 8, 10, 49, 51, 53, 54, 98, 100, 142, 215–218, 221, 250, 252, 255–256
- good
  - common, 3, 6–7, 11, 50, 98–99, 143, 171, 178, 220–222
  - customary (non-scientific) meaning of, 147–148, 152–155, 164–165
  - instrumental, 3, 7, 20, 103, 107, 156, 159, 165–178, 220
  - scientific (reformed) meaning of, 20, 42, 147–148, 153–155, 164–179
  - theory of
    - desire-fulfilment, 11, 40, 107–108, 143–145, 159–160, 163
    - hedonist. *See* hedonism
    - true (*bonum verum*), 43, 47–48, 155–156
  - ultimate / supreme (*summum bonum*), 1, 20, 98, 117, 139–143, 161–163
- Grotius, Hugo, 1–7, 11, 127, 130, 181, 199, 205, 207, 214, 276
- Hacking, Ian, 128–129
- hedonism, 11, 20, 107–108, 139–141, 143–145, 156–160, 161–164, *See also* Epicurus; Cyrenaics
  - psychological, 117, 146–147, 229
- honestas*, 1–4, 185, 206
- hope, 133–135, 136, 144, 150–151, 158–161, 162–163
- imagination, 30–33, 65–70, 74–76, 84–85, 87–88, 90, 151, 248, 267
  - pleasure of. *See* pleasure of anticipation
- imperatives, 55–57, 58–59, 91, 164
- intentionalism, 32, 57, 69, 73
- interpretation, legal, 237–241
- justice, 2–4, 6, 21–22, 49, 188, 207, 217–218, 226–227
- Lloyd, S. A., 224–226
- logic, 49–50, 51, 86, 115–116, 271–273, 275
- Lucretius, 159, 218
- materialism, 7, 27, 41, 65, 118, 263, 275
- meaning. *See also* definitions, reforming
  - customary (non-scientific), 17, 38, 42, 142, 148, *See also* good, customary meaning of
- morality / moral, meaning of, 22, 27, 219–223
- Murphy, Mark, 104–105, 132
- names
  - abstract, 81–82, 89
  - universal, 80–81, 82, 89
- naturalism, metaethical. *See* reductionism, ethical-naturalist
- natural-law theory, classical, 3, 97, 181, 191, 193
- noncognitivism. *See* prescriptivism
- obligation
  - directed, 22, 191–200, 228–230, 235, 258
  - eudaimonistic, 2, 4, 5, 6, 7, 9, 11, 181, 191, 193, 244
  - interest theory of direction, 195–196
  - juridical, 1–5, 6–7, 9, 21, 22, 61, 181, 191, 192, 193, 219, 228–229, 244, 261, *See also* obligation, directed
  - moral, 1, 2, 11, 22, 218–219, 223, *See* morality / moral, meaning of
  - natural, 5, 193, 215
  - will theory of direction, 196–199
- passions
  - post-mortem, 106–108, 131–132, 164
  - reactive, 210, *See also* attitudes, reactive; revengefulness
  - self-destructive, 101, 104–105, 106–107, 131–135
  - suicidal, 132–135
- person / personhood, 9, 23, 196, 215–216, 228, 245–260, 274, 276
  - artificial, 196, 216, 249–252, 255–259, 260
  - natural, 10, 196, 216, 248, 249–255, 257, 260
  - qua* representer vs. *qua* represented, 249–257
- Plato, 148, 205, 207
- pleasure, 20, 33–34, 70, 107–108, 139–141, 143–164
  - hybrid, 151–154
  - katastematic, 149–150
  - kinetic, 149–150
  - life of. *See* hedonism; felicity
  - of anticipation (imagination) / mind, 20, 108, 132, 141, 144, 148–154, 157–164, 170, 171, 206
  - of satisfaction (fruition) / sense, 20, 131, 141, 143, 154, 158, 159–160, 163, 170, 171
  - short-run, 47, 117
- positivism, legal, 235–237
- praise, 52–56, 58, 153, 164, 183, 215, 219

- Precepts  
 Juridical, 228, 274, 276  
 Prudential. *See* prudence, precepts of  
 prescriptivism, 12, 18, 29, 57–59  
 probability, 128–131  
 projectivism, 17–18, 29–30, 32–35, 36–39, 57, 267–269
- properties  
 artificial, 23, 251, 253, 263–267, 269–271, 273–274  
 evaluative, 30, 33–34  
 irrational, 263, 266–269, 273–274  
 natural, 8, 18, 40, 253, 263–266, 270–274  
 normative, 8, 17–18, 23, 27–29, 34–35, 39, 52, 95, 253, 257, 263, 273–275  
 real, 7–9, 23, 263–269, 273–274
- proposition  
 assenting to a, 56, 60, 72–76, 82–87, 89, 272–273  
 unity of the, 80, 85
- prudence, 6, 13, 129–130, 207, 227, *See also* sins of imprudence  
 epistemic vs. affective/practical, 111–113, 125, 169  
 precepts of, 19, 22, 109–110, 117, 119, 137, 138, 224, 228, 274
- prudentialism, 11, 19, 110–111, 131–132, 135–138, 141, 165–166, 227, 228, 230
- punishment, 4, 21, 178, 185, 199, 205–212, 214–215, 224, 231, 237, 240
- rationality, 15–16, 48–50, 115
- reason, dictate (precept) of, 2, 43, 48–51, 52, 56, 60, 103, 114
- reasoning, 14, 18–19, 23, 47–52, 58, 61, 63–66, 71–91, 97–98, 113–118  
 linguistic, 19, 51, 64, 78–83, 113, 115, 248  
 non-linguistic, 64, 77–78, 248  
 responsive to, 18, 63–65, 72, 82–91, 212
- reasons  
 affective, 1, 14, 19, 44, 63, 100–106, 111, 121–124, 166, 168, 212  
 epistemic, 14, 19, 20, 44–45, 91, 92, 111, 114, 117, 122, 126, 168–178, 212  
 explanatory, 14, 42–46  
 moral. *See* morality / moral, meaning of  
 motivational, 14, 46, 92, 130, 229  
 normative (notion of), 14–17  
 practical, 1, 3, 5, 14, 45–46, 62, 63, 95, 100–106, 111, 114, 117, 121–124, 131, 166, 186, 212  
 responsive / answerable to, 18, 21, 63–65, 72–73, 82–91, 181–183, 185–189, 201, 203, 212, 213, 246, 248–249
- reductionism, ethical-naturalist, 8, 10, 12, 18, 21, 27–28, 39–52, 56, 99, 199–200, 276
- representation, 209, 216, 245–246, 249–260  
 by fiction, 255–257  
 mental, 30–39, 68–70, 274
- responsibility  
 accountability sense of. *See* accountability  
 attribution sense of, 9, 23, 188, 201, 203, 212, 216, 240, 246, 249, 252–255, 258  
 causal sense of, 184–185, 200, 203, 248, 249
- revenge, 100, 184–185, 201–210
- revengefulness, 201–208
- right of nature, 9, 60, 119, 124, 127, 137, 205, 211, 230–231, 242–245, 246
- rights  
 inalienable, 233  
 to a life worth living, 22, 232  
 to bare survival, 22, 231, 234–235  
 perfect, 4–5, 199
- scholasticism, 1, 30–31, 32–33, 62, 67, 74, 90, 97, 183, 186, 213, *See also* Aquinas; Bramhall; Suárez
- self-preservation  
 meaning of, 135–138  
 non-social means of, 167, 168, 179, 220–221, 222  
 social means of, 6, 20, 22, 35, 167, 168, 172–178, 220, 222, 224, 226
- Seneca, 205–208
- sins of imprudence vs of injustice, 187–189, 217
- Skinner, Quentin, 116, 250–251
- sociability, natural, 3, 4–5, 7, 98–99
- Stoics, 1, 139, 140, 142, 167–168, *See also* Seneca  
 Suárez, Francisco, 1–5, 11, 13, 62, 181, 183, 219, 276
- subjectivism  
 about reasons, 8, 95–99  
 cognitive, 19, 119–131, 138  
 conative, 19, 99–105, 138  
 about value, 8, 40, 95, *See also* good, theory of, desire-fulfilment
- suicide, 131–135
- summum bonum*. *See* good, ultimate; felicity
- teleology, 7, 27, 97–99
- temperance, 109, 122, 167–168, 174, 179, 221, 222
- tranquility (*ataraxia*), 140–141, 149, 162
- transmission, instrumental, 19, 101–104, 125, 138, 261, 274
- truth, self-evident, 97–99, 116, 118–119, 275
- Tuck, Richard, 30, 36, 37, 128, 130, 199, 243
- utilitas*, 2–3, 36, 146, 206, 221
- well-being. *See* felicity