Punishment, Self-Defence, and Madness in Hobbes's *Leviathan*

John A. Petrakis

A Thesis Submitted to McGill University

in Partial Fulfilment of the Requirements

of the Degree of Master of Arts

Department of Political Science

McGill University, Montreal

April 2019

Table of Contents

Abstract / Résumé

This thesis delineates and resolves various puzzles which the right to self-defence ("RSD") introduces into the conception of punishment Hobbes presents in *Leviathan*. I first examine the *origin* of punishment. Upon presenting three central concepts in Hobbes's legal theory (personality, authorisation, and the alienation of rights), I demonstrate that the right to punish amounts to the sovereign's natural right exercised pursuant to the authorisation of each and every subject. Every subject authorises the punishment of every subject – even himself – thus inscribing punishment into the framework of public law. Next, I examine the *execution* of punishment. When combined with Hobbes's theory of authorisation, the retention of the RSD by every subject produces the paradox of the criminal with two contradictory wills: an authorised will to punish himself and a personal will to resist punishment. I develop an original account of madness in Hobbes and identify problems when examining the criminal through its lens. To conclude, I propose to resolve these problems by reconstructing the RSD. I argue that the rationale Hobbes provides for this right is the avoidance of death and extreme suffering, even though he claims the right can be opposed to all forms of punishment. If the scope of the RSD were narrowed in accordance with its underlying rationale, the RSD could not be opposed to most forms of punishment and the paradox of the two-willed criminal would be resolved.

Cette thèse cerne et résout divers problèmes que le droit d'autodéfense (DAD) introduit dans la notion de la punition que Hobbes introduit dans *Léviathan*. J'examine premièrement l'*origine* de la punition. Ayant présenté trois concepts centraux de la théorie légale hobbesienne (personnalité, autorisation et aliénation), je démontre que le droit de punir équivaut au droit naturel du souverain exercé selon l'autorisation de chaque sujet. Chaque sujet autorise la punition de chaque sujet – même sa propre punition – inscrivant donc la punition dans le cadre du droit public. En deuxième lieu, j'examine l'*exécution* de la punition. Lorsque combinée avec la théorie hobbesienne de l'autorisation, la rétention du DAD produit le paradoxe du criminel avec deux volontés contradictoires : la volonté autorisée de se punir ainsi que la volonté personnelle de résister sa punition. Je développe une théorie originale de la folie dans Hobbes et j'identifie des problèmes en l'appliquant au criminel. Pour conclure, je propose de résoudre ces problèmes en reconstruisant le DAD. Je soutiens que la justification fournie par Hobbes pour le DAD se rapporte à l'évitement de la mort et de la souffrance extrême, malgré les propos de Hobbes selon lesquels le DAD s'appliquerait à toute forme de punition. Si nous restreignons la portée du DAD en accord avec sa justification, le DAD ne s'opposerait plus à la plupart des punitions et le paradoxe du criminel avec deux volontés serait résolu.

Acknowledgements

This thesis owes its existence to my supervisor, Arash Abizadeh, who introduced me to the complexity of Hobbes's philosophy, encouraged me to explore these themes, and provided insightful comments at every stage of drafting. I am also grateful to Victor Muñiz-Fraticelli for his invaluable mentoring throughout my time at McGill.

I thank my family, and especially my father Petro, for supporting me during my studies. I dedicate this thesis to my grandfather Gregory and to the memory of my grandfather John.

Introduction

In *Leviathan*, Thomas Hobbes makes three well-known claims. First, he posits that human beings are motivated chiefly by self-interest. Second, that they hold an inalienable right to self-defence ("RSD"). Finally, that they can authorise a political ruler, whom Hobbes calls the sovereign, to act on their behalf through a social contract. At first glance, these claims work together: since our self-regarding desires dictate our behaviour and since we aim to defend ourselves, we should agree to enter a political community, as that is the surest way to protect our lives.[1] Moreover, once we find ourselves under the jurisdiction of a sovereign, we should obey his commands without difficulty, seeing as we have authorised them.

But Hobbes tells no such story. To the contrary, he is preoccupied with politically dangerous behaviour and devotes much of *Leviathan* to discussion of rebellion and criminality. Indeed, closer inspection of the mechanism he proposes for addressing these threats – *punishment* – reveals a tension within his political theory, as his RSD seems to cut against his idea of authorisation. Hobbes famously calls the subject who breaks the law because he believes it to be advantageous a "Foole",[2] and yet he insists that a subject who has already broken the law should resist the sovereign's attempts to bring him to justice. In the first case, the subject cannot appeal to self-interest to violate the laws he has authorised; but in the second case, his self-interest founds a right to resist the punishment for which he is eligible under similarly authorised

[1] Hobbes, *Leviathan*, chapter 17, paragraph 1, page 254/85. Future references to *Leviathan* are given as follows: *L*, chapter.paragraph, page/page. The page numbers preceding the / are to the 2012 Clarendon edition prepared by Noel Malcolm; those following the / are to the Head edition and are marked in Malcolm's text. When quoting passages from the secondary literature, I omit internal citations.
[2] *L*, 15.4-7, 222-24/72-73.

laws. Hobbes's vigorous defence of this right strikes a dissonant note within his overarching plea for political obedience.

In this thesis, I intend to delineate and resolve, at least partially, various puzzles which the RSD introduces into Hobbes's account of punishment.[3] In chapter 1, I explore the *origin* of the sovereign's right to punish. Hobbes appears to ground this right in two mutually exclusive sources. On the one hand, the social contract grants powers to the sovereign through the authorisation of his subjects, and Hobbes explicitly lists punishment among these powers.[4] On the other hand, Hobbes asserts that the right to punish is "not grounded on any concession, or gift of the Subjects", and that it finds its "foundation" in the sovereign's natural right.[5] He also asserts that no one can alienate his RSD.[6] As I explain later, many scholars assume that authorisation requires the transfer of a right; they proceed to infer that, since the RSD cannot be transferred, the sovereign cannot acquire the right to punish through authorisation. On this view, punishment is reducible to the exercise of the sovereign's natural right.

The question of whether punishment is authorised is crucial to Hobbes's political theory. A negative answer would fatally undermine his project of establishing a juridical relationship between subjects and their sovereign. Should authorisation play no role in punishment, the practice would amount to a conflict between the natural rights held by the sovereign and the criminal, necessarily occurring outside the boundaries of civil society. In fact, civil society could

[3] In the interest of conducting a manageable study, I draw exclusively on the 1651 *Leviathan*, which is widely regarded as Hobbes's most significant contribution to political thought. However, I reference some scholarship which uses textual evidence from his other works.
[4] *L*, 18.1, 264/88 and 18.4, 276/92.
[5] *L*, 28.2, 486/161.
[6] *L*, 14.8, 202/65-66.

hardly be said to exist if natural violence regulated its internal relations; Hobbes says as much when he distinguishes slaves from servants.[7] A commentator who espouses this view has eloquently stated the problem it entails: "Hobbes fails to discover and exhibit a right to punish – therewith also, of course, failing to give us a sovereign and thus a commonwealth."[8]

To elucidate the origin of Hobbesian punishment, I first expound three central concepts in Hobbes's legal theory: personality, authorisation, and the alienation of rights. I demonstrate that the social contract involves two distinct steps: the subjects, through mutual covenants, both *authorise* the sovereign to act in their name and also *alienate* a portion of their natural right as a gift to the sovereign. The independence of authorisation from alienation means that the subjects' retention of the RSD cannot represent a conceptual obstacle to authorised punishment. Next, I contend that the tradition misreads the problematic passages as addressing an antinomy between the sovereign's right to punish and his subjects' RSD. Hobbes is actually addressing a different question: if the right to punish is political, how could individuals in the pre-political state of nature transfer it to the sovereign? His answer is that prospective subjects do *not* transfer it: rather, they authorise the sovereign to use *his own* natural right to inflict violence on lawbreakers. This results in a public conception of punishment in concord with Hobbes's broader political commitments.

In chapter 2, I turn to the various problems the RSD generates with respect to the *execution* of punishment. To begin, this right imperils the stability, and even the possibility, of civil society. Hobbes ostensibly grants a wide scope to the RSD, thereby legitimising unlimited

[7] *L*, 20.10-13, 312-14/103-04.
[8] Schrock, 857.

disobedience after a single act of illegitimate disobedience. This undermines the permanence of political obligation: almost every subject will commit some offence at some time which attracts some kind of punishment, and if he were to take Hobbes's advice, he would confront the sovereign in an escalating spiral of violence. Moreover, if people in the state of nature shared this view, it is unlikely that they would ever agree to enter the social contract. They would predict that they would eventually have to wage war against a powerful sovereign instead of dealing with disorganised rivals in an anarchic system; the prospective costs might appear to outweigh the temporary benefits, and Hobbes's political enterprise would lose its rationalistic justification.

The coexistence of legitimate punishment with legitimate resistance also raises an interesting puzzle about Hobbes's conception of rationality. His theory of authorisation holds that the "words and actions" of artificial persons, such as the sovereign, are "*Owned* by those whom they represent".[9] Since punishment is authorised, the criminal "owns" the sovereign's will to punish him. And yet, Hobbes contends that the criminal should form a personal will to resist any punitive actions: thus, the criminal concurrently owns *two contradictory wills*. Hobbes's comments on contradiction render this predicament even more incongruous. He says that it is absurd to signify two contradictory things and that absurdity "may also be numbred amongst the sorts of Madnesse".[10] One would expect Hobbes to then treat the criminal as somehow "mad" or irrational, and yet his advocacy for the RSD causes him to veer in the opposite direction. In this manner, the RSD introduces tension between Hobbes's political theory and his theory of rationality, threatening the unity of the philosophical edifice erected in *Leviathan*.

[9] *L*, 16.4, 244/81.
[10] *L*, 14.7, 200-02/65 and 8.27, 122/39.

To address these issues, I begin by delimiting the scope of the RSD. I conclude that it is an individual right which entitles its holders to resist *most forms of punishment*, not only those which directly threaten their lives. Although not as capacious as that found in some revisionist scholarship, this version of the RSD suffices to create the aforementioned problems. To explore their repercussions for Hobbes's theory of rationality, I develop an original account of his theory of mental illness (or "Madnesse", in his terminology). I argue that reason should play an *active* and *participative* role in deliberation, and that Hobbes regards as mad four classes of people who deliberate unsoundly: (1) those with passions of abnormal intensity; (2) those with passions with abnormal content; (3) those with unguided passions; and (4) those with unguided or absurd thoughts. I then diagnose the criminal, showing that he is "mad" according to this theory when he breaks the law but *not* when he resists punishment. Happily, this inconsistency is not produced by Hobbes's deep assumptions about human nature; rather, it can be traced to the scope of his RSD.

To complete the second chapter, I advance that only a right to resist death and extreme suffering is consistent with Hobbes's thought. His explicit extension of the RSD to additional punishments rests on historically contingent assumptions that they would likely entail death or extreme suffering. I propose to reconstruct Hobbes's theory so as to restrict the RSD accordingly; criminals would then lose the right to resist punishment in any commonwealth which prohibits capital punishment and torture. In such circumstances, the will to resist punishment could only arise from defective deliberation – and both the lawbreaker and the fugitive would match Hobbes's account of madness. Concomitantly, submission to punishment

becomes the best response for any subject who has broken the law due to a temporary bout of madness – and knowledge of this fact renders the social contract attractive to rational agents.

1) The Origin of Punishment

Recent work on Hobbes recognises that "[the] structure and texture of his thought is densely juridical, his theory presented 'in familiar terms – "reason", "right", "law", authority", "obligation" – whose resonance is primarily legal'."[11] One need not go so far as to interpret Hobbes as a robust advocate for the rule of law, as some commentators have,[12] to appreciate the centrality of legal order to his philosophy. To remedy the human proclivity for violence, which threatens to turn "the life of man, solitary, poore, nasty, brutish, and short",[13] Hobbes prescribes the marriage of power and law, declaring that power should ordinarily (but not exclusively)[14] be exercised through legal forms and procedures.

Even this portrayal of Hobbes as a weak rule of law theorist – *i.e.* as someone who believes the relationship between the state and the individuals under its authority ought to be habitually, even if not always, mediated by law – is jeopardised by uncertainty about the origin of punishment in his political system. As I mentioned in the introduction and reiterate below, a significant academic current holds that Hobbesian punishment is *not* grounded in the social

[11] Poole, 68, citing Cromartie, 21.
[12] Over the past two decades, David Dyzenhaus has developed such a reading of Hobbes by underscoring the limitations placed by natural law on sovereign power and also the constitutional function of the judiciary in his political system: see Dyzenhaus (2001, 2010, 2012). In a similar vein, Fox-Decent conceptualises the Hobbesian sovereign's authority as a rule-bound, fiduciary relationship with his subjects.
[13] *L*, 13.9, 192/62.
[14] Poole advances a nuanced reading of Hobbes according to which the sovereign ordinarily makes his judgements known by enacting and enforcing public laws, while reserving the prerogative power to act outside the law to protect his subjects in exceptional circumstances.

contract; rather, it represents the sovereign's infliction of natural violence upon those who repudiate civil society. But if lawbreaking engendered a relapse to the state of nature, any juridical bond between sovereign and subject would be ephemeral: it could only last as long as the former believed the latter were obeying the law. Thus, the law could *never* resolve disputes between sovereign and subject. Beneath its veneer, there would only lurk brute force holding society together. On this account, "[the] whole structure of *Leviathan* is shaken…for the state of nature has never truly been transcended."[15]

In this chapter, I show that this unfortunate conclusion can be avoided by salvaging a public notion of punishment from Hobbes's text. First, I explicate three critical concepts in Hobbes's legal theory – personality, authorisation, and alienation – and the manner in which they interact during the creation of the social contract. Next, I demonstrate that the sovereign's right to punish each and every subject *does* flow from the consent of each and every subject. Although difficulties will remain to be addressed in chapter 2, I will have inscribed Hobbesian punishment within the framework of public law.

a) Personality, Authorisation, and Alienation

Generally speaking, jurists in early modern Europe revered the Roman civil law codified in the *Corpus iuris civilis*, considering it far more comprehensive and rational than native legal systems. While Hobbes cautions his readers that Roman law does not form part of the "civil law" of a commonwealth absent its explicit promulgation by the sovereign, it is evident that "Roman law nevertheless function[s] as a rich fund of concepts in framing central aspects of his civil

[15] Norrie, 308.

science."[16] Hobbes's familiarity with this tradition becomes apparent at the outset of chapter 16, where he defines the term "person". Despite borrowing liberally from the natural sciences in earlier chapters, at this stage Hobbes makes no mention at all of biology. He is not concerned with bodies *per se* but with the words and actions they produce – and, more importantly, with the rules that assign responsibility for those words and actions. Hobbes's theory of personality is really a theory of attributed action; as such, it constitutes the bedrock of two legal concepts – authorisation and alienation – deployed in his account of the social contract.

Hobbes opens chapter 16 as follows:

> 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.[17]

As David Runciman observes, three distinctions are woven into "this complex and subtle definition".[18]

1. Persons versus Non-Persons. A person is an entity capable of speech or action. Although Hobbes never uses the term "non-person", he implies that entities which cannot speak or act do not count as persons. As we will see, this story is complicated by the fact that such entities may nonetheless have words or actions attributed to them. Thus, a non-person has the latent capacity to become a person through attribution; "[but] what Hobbes does not

[16] Lee, 212; see 212-20 for a summary of the intellectual reception of Roman law in late medieval and early modern Europe. Hobbes's relevant remarks are at *L*, 26.1, 414/136-37.
[17] *L*, 16.1-2, 244/80.
[18] Runciman, 269-70.

say is that such capacity is always realised, and thus it follows that where unrealised, the thing in question is not a person."[19]

2. Natural versus Artificial Persons. A natural person represents himself, whereas an artificial person represents some other person.
3. True versus Fictional Representation. Within the category of "artificial" persons, Hobbes differentiates those who represent others "truly" from those who do so "by fiction".

Some clarifications are in order. To begin, a Hobbesian "natural person" matches the most common idea of a person: namely, a human being acting on his own behalf. However, some human beings are not natural persons because they lack reason and therefore cannot be held responsible for their words and deeds; such are "Children, Fooles, and Mad-men".[20] Furthermore, human beings endowed with reason are not always *natural* persons. Immediately following his definition of a person, Hobbes draws an analogy to the theatre and says that "a *Person*, is the same that an *Actor* is"; the relevant question is who the actor represents.[21] This goes to the second distinction: an actor is a natural person when he represents himself, but he is an "artificial person" whenever and to whatever extent he represents someone else.

Usually, an artificial person represents a natural person who commissioned him. Moving to the third distinction, this amounts to "true" representation in Hobbes's schema. As he writes,

> Of Persons Artificiall, some have their words and actions *Owned* by those whom they represent. And then the Person is the *Actor*; and he that owneth his words and actions, is the AUTHOR: In which case the Actor acteth by Authority. For that which in speaking of goods and possessions, is called an *Owner*, and in latine *Dominus*, in Greeke κυριος, speaking of Actions, is called Author. And as the Right of possession, is called Dominion; so the Right of doing any Action, is called AUTHORITY. So

[19] *Ibid.*, 270.
[20] *L*, 16.10, 248/82.
[21] *L*, 16.3, 244/80.

> that by Authority, is alwayes understood a Right of doing any act: and *done by Authority*, done by Commission, or Licence from him whose right it is.[22]

This describes a straightforward principal-agent relationship, where the principal is called an "author" and the agent is called both an "artificial person" and an "actor". Despite the clear meaning of the text, Quentin Skinner argues that it makes better sense to treat the *principal* as the "artificial person"; he tenders excerpts from Hobbes's later works as evidence that he eventually adopted this position.[23] Although this terminological issue has no real bearing on my arguments, I wish to clarify that I will employ Hobbes's terms as they are defined in *Leviathan*. In truth, I share Runciman's view that

> the English *Leviathan*…is where [Hobbes] gives his fullest account of the concept of the 'person'. The latter versions represent a simplification rather than an embellishment of this earlier account, and in some respects a retreat from the difficult issues it raises.[24]

Moreover, with respect to the final category (representation "by fiction"), Skinner's interpretation creates unnecessary inconsistency between Hobbes's model and the Roman law doctrines to which he alludes. Representation by fiction refers to a situation where the words and deeds of the actor are attributed to an entity which was incapable of authorising them. While this might seem strange at first, Hobbes offers several commonplace examples. For instance,

> Inanimate things, as a Church, an Hospital, a Bridge, may be personated by a Rector, Master, or Overseer [even though] things Inanimate, cannot be Authors, nor therefore give authority to their Actors: Yet the Actors may have Authority to procure their maintenance, given them by those that are Owners, or Governours of those things. And therefore, such things cannot be Personated, before there be some state of Civill Government.[25]

[22] *L*, 16.4, 244/81.
[23] Skinner, 187-96.
[24] Runciman, 268.
[25] *L*, 16.9, 246/81-82 (emphasis mine).

Similarly, children and insane adults "may be Personated by Guardians, or Curators; but can be no Authors" as long as they lack reason.[26]

In these instances, "[r]epresentation by fiction is the representation of what are otherwise non-persons… The fiction is that they truly are persons, truly capable of the actions that personal responsibility requires."[27] Hobbes stresses that this kind of representation is premised upon a system of civil laws which assigns ownership or governance of non-persons to certain natural persons. In turn, the "owner" of a non-person can authorise a third party to represent that same non-person. The classic example is the guardianship of a child: the law renders the parents (or the father, in Hobbes's day) responsible for the child, and they may appoint a guardian to act on the child's behalf in legal proceedings. Hobbes broke no new ground here, as Roman authors proposed this very mechanism for allowing incapacitated principals to exercise legal rights. As Daniel Lee explains,

> while the underage or mentally incompetent ward was technically regarded a *dominus* [*i.e.* principal], as the passive beneficiary and author of the guardian's actions, it was the guardian that conferred a jural personality on the ward and enabled the ward, by legal fiction, to be an author…[28]

Skinner's reading unsettles Hobbes's evident affinity to the Roman jurists without any obvious advantage. For Skinner, the capacity *to be represented* is the condition of personality, and so all kinds of entities – even those incapable of speech and action – *always* count as persons. Since he believes the term "artificial person" attaches to the principal, Skinner forces three kinds of entities into this category: natural persons who directly authorise an agent; "purely artificial persons" who can be "truly" represented, such as children and buildings; and "wholly

[26] *L*, 16.10, 248/82.
[27] Runciman, 272.
[28] Lee, 225-26.

fictitious" persons, such as characters in a play.[29] This alternative model, which elides the concept of representation by fiction, has two drawbacks. First, it invites confusion by blurring distinctions drawn by Hobbes. Second, it masks Hobbes's debt to Roman jurisprudence, which maintained that mentally incompetent human beings did *not* hold legal personality *unless and until* an agent were appointed to act on their behalf. The existence of the agent conferred personality upon the principal, in stark contrast to a normal agency contract. The theory presented in *Leviathan* is attentive to this difference and accommodates it.[30]

This discussion has revealed that Hobbes's theory of personality is entwined with a theory of *authorisation*. He is most interested in persons who represent others through a grant of authority, for an obvious reason: his social contract is essentially an elaborate scheme of authorisation. According to Hobbes, private judgement generates conflict in the state of nature;[31] to attain peace, individuals must authorise a sovereign to act and will in their place, at least with regard to political questions. A subject is understood to own the actions and judgements of the sovereign, even though they proceed from the sovereign's body and mind. Hobbes makes this point rather dramatically, proclaiming that persons who institute a sovereign

> submit their Wills, every one to [the sovereign's] Will, and their Judgements, to his Judgment... [A]s if every man should say to every man, *I Authorise and give up my Right of*

[29] Skinner, 191-95.
[30] Abizadeh (2018), 245-62, develops an original account of Hobbesian personality which bridges the disagreement between Skinner and Runciman. According to Abizadeh, in an agency relationship, the representer and representee both represent the *same* artificial person; thus, Hobbes expressed a consistent conception of personality in his various works. For example, if I were to hire a lawyer to represent me in court, the lawyer would "be" (as representer) the same artificial person I "am" (as representee). It is beyond my remit to assess the validity of this complex interpretation. However, it does not appear to contradict the point I am making here. With respect to representation by fiction, Abizadeh does not suggest (as Skinner does) that an entity which is not a rational human being counts as a person outside a relationship of representation.
[31] *L*, 5.3, 66/18-19.

> *Governing my selfe, to this Man, or to this Assembly of men, on this condition, that thou give up thy Right to him, and Authorise all his Actions in like manner*.[32]

However, as this passage shows, authorisation might not be the only legal mechanism at play in the social contract. The question is whether authorisation necessarily implies the alienation of a right. Stated differently, does the phrase "I Authorise and give up my Right of Governing my selfe" refer to a single legal act or two distinguishable ones? To answer this query, I will first canvass Hobbes's concept of the alienation of rights; afterwards, I will return to authorisation and show that it is an independent concept. I will then be in a position to explain the interaction of both concepts in the social contract.

Throughout *Leviathan*, Hobbes adheres to the classical Roman understanding of rights and obligations. Indeed, much of chapter 14 can be read as an extended gloss on the famous maxim *obligatio est iuris vinculum*.[33] Hobbes equates rights with liberties, defined as "the absence of externall Impediments" to action.[34] Thus, several persons might have a right to the same thing and they might choose to resolve the matter through violence, each becoming an "impediment" to the others. However, they might also make a different choice: one or more amongst them might *alienate* their right to the thing under dispute. As Hobbes says, "To *lay downe* 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."[35] In other words, one person agrees to stand out of the way of others when they exercise their own rights.

---

[32] *L*, 17.13, 260/87 (emphasis mine, italics in the original).
[33] "An obligation is a legal bond [or chain]." The maxim is found in Justinian's *Institutes*: book 3, title 13, preamble. See also Dedek and Schermaier.
[34] *L*, 14.2, 198/64.
[35] *L*, 14.6, 200/65.

Hobbes specifies that this process can take one of two forms, *renunciation* or *transfer*:

> Right is layd aside, either by simply Renouncing it; or by Transferring it to another. By *Simply* RENOUNCING; when he cares not to whom the benefit thereof redoundeth. By TRANSFERRING; when he intendeth the benefit thereof to some certain person, or persons.

Moreover, there are two ways to transfer a right. A *contract* involves "[t]he mutuall transferring of Right",[36] whereas a *gift* involves "the transferring of Right [which] is not mutuall".[37] Ultimately, every species of alienation (renunciation, contract, and gift) creates a corresponding obligation not to hinder others within the scope of the abandoned right – Hobbes defines any such hindrance as "Injustice".[38]

Nevertheless, Hobbes warns that some alienations cannot be valid:

> Whensoever a man Transferreth his Right, or Renounceth it; it is either in consideration of some Right reciprocally transferred to himselfe; or for some other good he hopeth for thereby. For it is a voluntary act: and of the voluntary acts of every man, the object is some *Good to himselfe*. And therefore there be some Rights, which no man can be understood, by any words, or other signes, to have abandoned, or transferred.[39]

In this passage, Hobbes explicitly evokes the common law doctrine of "consideration" but does not actually endorse it. This doctrine holds that a contract is enforceable only if the parties reciprocally promise to exchange something with economic value, and thereby excludes gifts. In contrast, Hobbes acknowledges that someone might give away a right not because he expects a tangible benefit, but rather

> in hope to gain thereby friendship…or in hope to gain the reputation of Charity, or Magnanimity; or to deliver his mind from the pain of compassion; or in hope of reward in heaven.[40]

---

[36] *L*, 14.9, 204/66. Hobbes adds that a "covenant" is a kind of contract where one of the parties immediately transfers the right to a thing but delivers the thing itself at a later time: *L*, 14.10-11, 204/66.
[37] *L*, 14.12, 204/66.
[38] *L*, 14.7, 200-02/65.
[39] *L*, 14.8, 202/65-66.
[40] *L*, 14.12, 204/66.

Hobbes's further contention that the promise of a gift creates binding obligations, at least in certain circumstances,[41] is incompatible with the doctrine of consideration as understood either in his day or ours, but perfectly compatible with the less stringent doctrine of *causa* developed by the medieval glossators of Roman law.[42]

Thus, Hobbes's insistence that some alienations are invalid is not intended to exclude gifts motivated by goodwill or the mere *hope* of material benefit. In fact, he is not concerned with the motives behind an alienation but rather with its effects. He focuses on rights whose abandonment can *only* cause harm, such as the right to defend one's life, to resist imprisonment, or to make use of "food, ayre, [or] medicine".[43] In essence, these are all iterations of the RSD: a bundle of rights contained within, but not exhaustive of, our broader natural right to do anything we deem conducive to self-preservation.[44] While I review the scope of this right in chapter 2, the important point for present purposes is that Hobbes believes we can only hurt ourselves by laying down certain rights; for this reason, we should never be understood to do so.

We can now return to the concept of authorisation and investigate its relationship to the alienation of rights. When he sets out the legal consequences of authorisation, Hobbes implies that the author *does* sometimes surrender a right to the actor. For instance, a person who authorises an agent to conclude a contract in his name with a third party is thereafter bound to

---

[41] *L*, 14.15, 206/67.
[42] The doctrine of *causa*, which survives in civil law jurisdictions, holds that any promise made with a serious intention, or "cause", creates an enforceable contractual obligation. This captures promises made with the expectation of economic gain but also those motivated by generosity or liberality. For excellent historical and analytical discussion of the doctrines of *causa* and consideration, see Gordley, 49-57, 137-39, and 164-75, and Zimmerman, 549-59.
[43] *L*, 14.8, 202/66 and 21.12, 336/111.
[44] *L*, 14.1, 198/64.

respect any such contract. However, if the actor were to conclude a contract outside the scope of his grant of authority, such a contract would not bind the author.[45] Indeed, "when the Authority is feigned, [the Covenant] obligeth the Actor onely; there being no Author but himselfe."[46] The implication is that the author transfers a right he previously held to the actor on the condition that the latter exercise it for him, a procedure commonly called the "extension" of a right.[47] Once we delimit the scope of the extended right, we can impute responsibility to the author for certain actions performed by the actor.

The next question is whether the extension of an existing right is *necessary* for valid authorisation. As I demonstrate in the following subchapter, many scholars answer in the affirmative and argue on that basis that Hobbes's social contract is defective. However, Hobbes contemplates two situations where the author instructs the actor to perform deeds *which the author had no right to perform himself*. Both passages strongly suggest that authorisation remains valid and carries legal consequences in these circumstances. Shortly after introducing the concept in chapter 16, Hobbes considers a situation, presumably in the state of nature,[48] where the actor has pledged to represent the author and is told to infringe the law of nature in some way. Hobbes places all responsibility for the breach upon the author:

> When the Actor doth any thing against the Law of Nature by command of the Author, if he be obliged by former Covenant to obey him, not he, but the Author breaketh the Law of Nature: for though the Action be against the Law of Nature; yet it is not his: but contrarily, to refuse to do it, is against the Law of Nature, that forbiddeth breach of Covenant.[49]

[45] *L*, 16.5-6, 246/81.
[46] *L*, 16.8, 246/81.
[47] Green, 28.
[48] Hobbes's hypothetical only mentions natural law, not civil law. Moreover, it occurs before his presentation of the social contract in chapter 17.
[49] *L*, 16.7, 246/81.

Clearly, the grant of authority is valid in this case. But is the actor performing deeds the author had no right to perform? One could dispute this on the basis that natural law does not truly constrain an individual's natural right to act as he pleases in Hobbes's philosophy, but rather amounts to prudential advice.[50]

Fortunately, we can leave this thorny issue aside because *Leviathan* contains another passage germane to our discussion which does not implicate natural law. In chapter 27, Hobbes turns to individuals who live in a commonwealth and are subject to its civil laws. He explains that an author cannot seek redress from an actor who broke the law in his name "because no man ought to accuse his own fact in another, that is but his instrument". However, he instantly clarifies that such lawbreaking "is not Excused against a third person thereby injured; because in the violation of the Law, both the Author, and Actor are Criminalls."[51] Here, the author most certainly did *not* have the right to break the law. Nevertheless, two legal consequences stand as evidence that the grant of authority was valid: (1) the author's inability to hold "his own fact" against the actor and (2) the author's liability for criminal punishment. If the grant of authority were invalid, the ascription of these consequences to the author could not be explained. Furthermore, the joint criminal liability of the actor poses no difficulty. The social contract binds every subject in a commonwealth to obey the law, and so it stands to reason that anyone implicated in lawbreaking is eligible for punishment. In fact, even in the passage from chapter

[50] Whether natural law obligations are binding or prudential might be the most intractable question in Hobbes scholarship. For leading statements of the competing positions, contrast Taylor and Warrender (they are binding) to Nagel (they are prudential); for a summary of the academic debate as it stood in the early 1970s, see Barry. Recently, Abizadeh (2018) has argued that Hobbes's ethical system contains two dimensions: (1) prudential "reasons of the good" and (2) obligatory "reasons of the right"; he places the laws of nature in the first category. At 228-35, Abizadeh discusses the relation between the laws of nature and obligations.

[51] *L*, 27.27, 470/157.

16, the actor avoided moral responsibility for his deed only because he was under a pre-existing natural law duty to obey the author. In a commonwealth, a subject can never owe a duty of unconditional obedience to any private person; in consequence, criminal responsibility extends both to the person who directly breaks the law and the one who instructed him to do so.

Philosophers of action commonly use the term "commission" to describe the mechanism whereby one person (the author) becomes responsible for the actions of another (the actor). Moreover, they recognise that a commission does not necessarily involve the extension of a right.[52] One such philosopher, David Copp, nevertheless claims that authorisation always involves an extension in Hobbes's system, so that an author can never commission an actor to do something he had no right to do himself.[53] The two preceding examples demonstrate that Copp is mistaken: Hobbes directly contemplates commissions without an extension.[54] In both Hobbes's theory and most legal systems, "the person who hires the hit man bears some of the responsibility of the murder"[55] – even though he had no right to murder anyone in the first place.

One final task remains for this subchapter: to determine the manner in which alienation and authorisation interrelate in Hobbes's social contract. I contend that Hobbes envisages two different, concurrent legal operations when he says that each party to the social contract postulates: "I Authorise and give up my Right of Governing my selfe…"[56] The first step – *authorisation* through mutual covenanting between individuals – is the most significant: it

[52] Copp, 589-93; Green, 26-32. Hobbes himself uses the word "Commission": see *L*, 16.4, 244/81, reproduced in the text corresponding to footnote 22.
[53] Copp, 585-89 and 593.
[54] *L*, 16.7, 246/81 and 27.27, 470/157.
[55] Green, 32.
[56] *L*, 17.13, 260/87.

creates the state, appoints the sovereign as the representative of the state, and grants the sovereign certain rights. The second step – *alienation* of certain entitlements by subjects to the sovereign – simply adds to the sovereign's arsenal of rights. I will now describe each step.

In the first instance, individuals *assemble* and *institute a state* (or, in the Hobbesian terminology, a commonwealth). As Skinner and Runciman observe, however, the state is a person which can only act through a representative (named "sovereign").[57] Hobbes is clear that the assembled individuals, by majority vote, can appoint a sovereign to represent the state "by fiction"; an interesting question is how they acquire the right to authorise this representation. Unlike the property owner who appoints an overseer for his bridge, these individuals do not obtain this right through an existing system of civil laws. Skinner suggests that the right is grounded in their relationship of "dominion" over the state, similar to that between a mother and her child.[58] Runciman is unconvinced, arguing instead that the members

> create the conditions which allow the actions of the sovereign to be attributed to them as a single unit, since they are jointly committed to taking responsibility for what the sovereign does. They thus make possible the fiction that they can act as a unit, and commit themselves to the real actions that can maintain that fiction. Unlike the bridge, the state does not exist at all before its representative is in place. But like the state, the bridge does not exist as a person without a representative...[59]

Arash Abizadeh offers a more convincing explanation than either of these authors.[60] He asserts that the social contract always begins by creating a democracy: the very act of *assembling* simultaneously creates the state *and* a sovereign democratic assembly which represents the state. Through their agreement to form a state, naturally equal individuals commit themselves to a

[57] Skinner, 196-204; Runciman, 272-73. Hobbes claims that a "Multitude of men" can only act through a designated representative: *L*, 16.13-14, 248-50/82.
[58] Skinner, 201-04.
[59] Runciman, 273.
[60] See Abizadeh (2016), especially 411-415.

majority-rule decision-making procedure; as such, they no longer constitute a disorganised "multitude". Abizadeh demonstrates that Hobbes makes this claim explicitly in *De Cive*, and he persuasively argues that the same claim is implicit in *Leviathan*.

Nevertheless, this primordial democracy need not endure. Hobbes is clear that the original assembly can transfer its sovereignty to an aristocratic assembly or an individual monarch by majority vote. Indeed, the rights of sovereignty are held by the person "on whom the sovereign power is conferred by the consent of the people assembled"; "every one, as well he that *Voted for it*, as he that *Voted against it*, shall *Authorise* all the Actions and Judgements, of that [sovereign], in the same manner, as if they were his own".[61] Thus, the new sovereign is authorised by every subject, even those who did not vote for him: it is one's prior consent to assemble which binds one to the social contract, *not* one's subsequent exercise of the original right to vote. The sovereign represents the state directly and every subject (as an individual who voluntarily incorporated himself into the state) at a remove.

This model grants the sovereign, whether an assembly or an individual, two kinds of rights which are critical to Hobbes's political project.[62] First, the sovereign acquires status rights as the sole representative of the state. Hobbes relies on these rights to rebut two alternative claims made by his contemporary opponents: that the people in some pre-institutional form can exercise sovereignty or that Parliament represents the people.[63] Second, the sovereign acquires immunity rights. The upshot of authorisation is that every subject owns the sovereign's actions;

[61] *L*, 18.1-2, 264/88.
[62] Green, 32-37. My writing assumes that the sovereign is an individual monarch, but the analysis would not change if it were an assembly (whether democratic or aristocratic).
[63] See Skinner, 204-08.

as a result, no subject may accuse the sovereign of injustice, since it is impossible to act unjustly against oneself.[64]

Despite its intricacy, this account of the social contract is still incomplete. As Michael Green shows, it is the procedure of *alienation* which grants the sovereign some other rights which are best characterised as exclusive liberties and powers.[65] Specifically, when they take on an obligation of obedience, the subjects agree not to interfere with the sovereign's exercise of his own natural right. They also surrender their natural right to govern themselves to him, thus grounding his power to make laws. This step is logically independent from authorisation, since it would be possible to alienate one's right to self-government to a tyrant (perhaps in exchange for periodic monetary payments) without authorising him and taking ownership of his actions.

Nevertheless, Hobbes maintains that there exists no contract whatever between a subject and his sovereign.[66] As I have shown, the social contract is really a network of covenants between subjects, which constitutes the state and commissions its sovereign representative. The individual subject does not promise *the sovereign* that he will obey his laws; rather, he promises *every other subject* that he will do so. In the absence of a contract with the sovereign, though, how is it possible for a subject to alienate certain rights to him? The answer is that these alienations are *gifts* which each subject makes to the sovereign's benefit in order to fulfil his

---

[64] *L*, 18.6, 270/90.
[65] Green, 32-37.
[66] *L*, 18.4, 266/89: "Because the Right of bearing the Person of them all, is given to him they make Soveraigne, by Covenant onely of one to another, and not of him to any of them… That he which is made Soveraigne maketh no Covenant with his Subjects before-hand, is manifest…" This applies to commonwealths instituted through a social contract, which have been the exclusive focus of this discussion. In *L*, 20.1-3, 306/101-02, Hobbes writes that in a "*Common-wealth by Acquisition*", *i.e.* one conquered in war, every conquered subject makes a contract with the victorious sovereign, trading his obedience for protection.

pledge to his co-subjects.[67] It is entirely possible for two or more persons to agree that they will each give an equivalent gift to a third party. Moreover, as previously explained, Hobbes considers the promise of a gift binding because he does not endorse the common law doctrine of consideration. He merely prohibits rights-transfers whose only possible effect is to inflict harm upon the transferor.[68] The multi-party gift envisaged in Hobbes's social contract is certainly not among their number; to the contrary, the expected outcome – social peace – is beneficial to every transferor and would likely even satisfy the requirements of the consideration doctrine.

Hence, Hobbes's social contract combines authorisation and alienation in an original and sophisticated manner. While certain scholars accuse Hobbes of inconsistency, they go astray by failing to grasp that authorisation is a commission which does not require the extension of a right.[69] As Green affirms, "[Hobbes's] social contract is internally consistent because it is possible to authorize someone else's actions by taking ownership of them while alienating one's own rights."[70] Now that we have a proper model of the social contract in view, we can trace specific rights held by the sovereign to specific combinations of authorisation and alienation. The right to punish is one such right, whose complex origin remains hotly debated.

---

[67] Abizadeh (2018), 197. I further observe that Hobbes says we owe a natural law duty of gratitude towards those who have graced us with a gift: *L*, 15.16, 230/75-76. Later, he claims that a sovereign who punishes innocent subjects is guilty of ingratitude: *L*, 28.22, 492/165. This implies that the sovereign receives liberty rights as gifts from his subjects.

[68] See footnotes 39-44 and the corresponding text.

[69] Green, 28-29, summarises this accusation, which has been made most famously by Jean Hampton and A. P. Martinich. In his most recent work on this subject, Martinich, 315-24, accepts the logical independence of alienation and authorisation. However, he continues to maintain that their specific combination in the social contract creates inconsistency. He interprets Hobbes as grounding sovereignty purely on authorisation in chapter 21 of *Leviathan*, and prefers this alleged model to the authorisation-alienation hybrid in chapter 17. Since the account of the social contract presented in this subchapter contains no inconsistency, there is no need to follow Martinich in reading an evolution in Hobbes's thought between these chapters.

[70] Green, 36.

b) Punishment as a Public Institution

The sovereign's right to punish his subjects forms a critical component of Hobbesian civil society. In chapter 28, Hobbes defines punishment as

> *an Evill inflicted by publique Authority, on him that hath done, or omitted that which is Judged by the same Authority to be a Transgression of the Law; to the end that the will of men may thereby the better be disposed to obedience*.[71]

Punishment is therefore the primary instrument by which the sovereign can compel his subjects to obey the law.[72] Since human beings enter civil society precisely to escape the evils of private judgement, through voluntary submission to a set of common standards of justice, the sovereign's right to enforce such standards is indispensable to the success of this enterprise – as is an adequate philosophical grounding for this right.[73]

When providing this grounding, Hobbes appeals to the notion of authorisation. In chapter 18, he offers an extensive list of the rights held by a sovereign instituted by individuals who agree to "*Authorise* all [his] Actions and Judgements". Among these is the right "of Punishing with corporall, or pecuniary, punishment, or with ignominy every Subject according to the Law…"[74] The suggestion that punishment is authorised by the subjects resonates with the definition reproduced above, which states that punishment is inflicted "by publique Authority".[75] Indeed, throughout chapter 28, Hobbes insists that the public character of punishment distinguishes this form of violence from various acts of "hostility". In particular, violence is

[71] *L*, 28.1, 482/161.
[72] *L*, 28.24-26, 494-96/166 points to another, positive method of encouraging compliance with the law and service to the commonwealth: the bestowing of rewards.
[73] For the problem of private judgement, see footnote 31 and the corresponding text. *L*, 26.3, 414/137 defines the civil law as "*those Rules, which the Common-wealth hath Commanded [every Subject], by Word, Writing, or other sufficient Sign of the Will, to make use of, for the Distinction of Right and Wrong*" – *i.e.* as a common standard of justice.
[74] *L*, 18.1, 264/88 and 18.14, 276/92.
[75] *L*, 28.1, 482/161.

hostile if it is inflicted without a public hearing, by a usurper, without a deterrent intention, in excess of the limits prescribed by law, or through the retroactive application of law.[76] These constraints track the contemporary notion of the rule of law and intimate that punishment is an official act that cannot be meted out on the sovereign's private whim.

Even more significantly, Hobbes differentiates the sovereign's suppression of rebels who repudiate civil society – a hostile act which "fals not under the name of Punishment", and which need not follow "the Punishments set down in the Law, [which are addressed] to Subjects, not to Enemies"[77] – from his treatment of ordinary criminals. He also says that a private revenge is not a punishment.[78] These efforts to entrench punishment within the juridical framework of the social contract are consistent with Hobbes's larger ambition: to persuade his readers that submission to a sovereign who punishes in conformity with law is preferable to the state of nature, where private persons inflict violence upon one another arbitrarily and unpredictably.

Despite all this textual evidence, a great number of scholars have expressed doubt that Hobbesian punishment is properly grounded in authorisation. Their starting point is a passage which immediately follows Hobbes's formal definition of punishment and casts a shadow over its purportedly "publique" character. This passage is worth reproducing at length:

> [There] is a question to be answered, of much importance; which is, by what door the Right, or Authority of Punishing in any case, came in. For by that which has been said before, no man is supposed bound by Covenant, not to resist violence; and consequently it cannot be intended, that he gave any right to another to lay violent hands upon his person. In the making of a Common-

[76] *L*, 28.5-7, 484/162 and 28.10-11, 486/162-63.
[77] *L*, 28.13, 486/16; see also *L*, 28.22-23, 492-94/165-66. In this regard, Hobbes stands in stark contrast to Jean-Jacques Rousseau, who views common criminals as enemies of the state: "every evil-doer who attacks social right becomes a rebel and a traitor to the fatherland by his crimes, by violating its laws he ceases to be a member of it, and even enters into war with it... when the guilty man is put to death, it is less as a Citizen than as an enemy..." (*Of the Social Contract*, book 2, chapter 5, paragraph 4).
[78] *L*, 28.3, 484/162.

> wealth, every man giveth away the right of defending another; but not of defending himselfe. [...] It is manifest therefore that the Right which the Common-wealth (that is, he, or they that represent it) hath to Punish, is not grounded on any concession, or gift of the subjects. But I have also shewed formerly, that before the Institution of Common-wealth, every man had a right to every thing, and to do whatsoever he thought necessary to his own preservation; subduing, hurting, or killing any man in order thereunto. And this is the foundation of that right of Punishing, which is exercised in every Common-wealth. For the Subjects did not give the Soveraign that right; but onely in laying down theirs, strengthened him to use his own, as he should see fit, for the preservation of them all: so that it was not given, but left to him, and to him onely...[79]

Hobbes incontrovertibly says two things here: (1) no subject can validly transfer his right to resist violence (*i.e.* his RSD) to the sovereign; and (2) the sovereign exercises his own natural right when he punishes any subject. Sceptical scholars add the following premise: (3) authorisation requires the valid transfer of a right from the author to the actor. Combining (1) and (3), they conclude that the sovereign's right to punish cannot be authorised. Turning to (2), they argue that punishment is an apolitical act inflicted by the sovereign on criminals who have relapsed to the state of nature. On this view, punishment expresses *de facto* authority without any legal foundation. As Jean Hampton puts it,

> The idea is that in order to give the sovereign the power to punish subjects, each person does not relinquish her right to defend herself, but only her right to all things, so that *in fact* the sovereign becomes powerful enough to inflict on any of them whatever punishments he decides are appropriate.[80]

This argument restores some conceptual harmony, even though it entails the paradoxical conclusion that the sovereign lacks the normative authority to fulfil his essential functions.[81]

To mitigate this difficulty, some interpreters who share premise (3) observe that it does not prevent a subject from authorising the punishment of every subject *except himself*. As I elaborate in chapter 2, the Hobbesian subject clearly grants the sovereign the major part of his

[79] *L*, 28.2, 482/161-62 (emphasis mine).
[80] Hampton, 191.
[81] Because he shares Hampton's assumptions, Schrock, 857, ominously announces that "Hobbes fails to discover and exhibit a right to punish – therewith also, of course, failing to give us a sovereign and thus a commonwealth." For his argument that punishment is not authorised by any subject, see especially 868-73.

natural right to inflict violence on others; he only retains its inalienable core, the RSD. Therefore, it is possible to claim that criminals revert to the state of nature *and* that the sovereign punishes them on behalf of his subjects, who validly gave him their pre-existing right to harm enemies. Nevertheless, the scholars who make this claim often acknowledge that it still places the right to punish in a unique and troubled position within Hobbes's political theory. Thus, David Gauthier remarks that "the right to punish is subtly different from all the other rights of the sovereign, in that the sovereign in each act of punishing is not exercising a right given him by all of his subjects." Alice Ristroph adds that "punishment is, at best, imperfectly legitimate" because it is not "universally and unequivocally authorized."[82]

This dissatisfaction is inescapable because the "partially authorised" account of punishment does no better than the "unauthorised" account at resolving the fundamental problems introduced by the notion that punished individuals stand in a relationship of lawless conflict with the sovereign. As I stated previously, Hobbes turns his mind to rebels who "having been subject to [the Law, are now] professing to be no longer so", and asserts that the sovereign does not engage in "punishing" when he acts against them.[83] To the contrary, the sovereign's actions amount to the discretionary exercise of his natural right and need not respect the safeguards worked into the definition of punishment. If criminals are in the same normative position as these rebels, then Hobbes's ideal of lawful punishment is an empty promise; it becomes impossible for a system of laws to ever regulate the relationship between individuals and the authorities they consensually appoint. As soon as a subject puts a foot wrong, and

[82] Gauthier, 148; Ristroph, 116. The claim that punishment is partially authorised is also made in Cohen, 38-44, and Steinberger, 861-63.
[83] See footnote 77 and the corresponding text.

perhaps even as soon as he is *suspected* of this, he is cast out of society and into the realm of arbitrary violence; whether or not that violence is authorised by his former confederates does nothing to change its extra-legal nature.

For this reason, a final group of scholars read Hobbes as advancing not a coherent theory of punishment, but rather two irreconcilable claims. In Susanne Sreedhar's words,

> [Hobbes] claims at one point that subjects are the authors of their own punishment and that they are punished by their own authority. [But] he suggests that the sovereign's right to punish cannot be attributed to his authorization in the social contract. When he must account for the sovereign's right to punish, he rests in squarely upon the sovereign's original right of nature.[84]

Alan Norrie concurs:

> Hobbes's resolution of the problem is unsatisfactory because the logic of his argument either leads to the conclusion that the individual does indeed grant the Sovereign the right to punish (despite what Hobbes claims) or leads to the immanent collapse of, and the implicit denial of the possibility of, the social state and the institution of punishment.[85]

Of course, all these problems would vanish if it were the case that the Hobbesian subject authorises the punishment of every subject *including himself*. The various scholars I have referenced deny this possibility because they hold premise (3), *i.e.* that authorisation requires the valid transfer of a right. However, I established the falsity of this premise in the preceding subchapter: for Hobbes, an authorisation is a commission which does not necessarily require a rights-transfer. The subject may authorise the sovereign to punish him while retaining the right to resist punishment. Although this might appear counterintuitive – and engenders problems of execution discussed in chapter 2 – it is in fact Hobbes's position. In two separate chapters, Hobbes affirms expressly that a person *can* covenant "*Unlesse I do so, or so, kill me*" or "*Kill*

[84] Sreedhar, 98.
[85] Norrie, 307; see generally 301-09.

me, or my fellow, if you please". What a person *cannot* covenant is "*Unlesse I do so, or so, I will not resist you, when you come to kill me*" or "*I will kill my selfe, or my fellow*."[86]

How does this account square with the passage stating that the "foundation" of punishment lies in the sovereign's natural right? Arthur Yates argues convincingly that the tradition misreads Hobbes as addressing a contradiction between the sovereign's right to punish and the subject's RSD in chapter 28. The puzzle which Hobbes actually identifies and resolves in the disputed passage is the following: How can the sovereign acquire a right which his subjects never possessed themselves? If punishment is inherently political, persons in the state of nature lack the right to punish others and are thereby incapable of transferring such a right to the sovereign. Hobbes responds that prospective subjects do not transfer this right – and that this poses no problem. The subjects simply authorise the sovereign to use his natural right to inflict violence on lawbreakers, while they concurrently surrender their own right to do the same. As Yates puts it, "The natural right to violence of the person (or persons) who holds the office of sovereignty is *laundered*, so to speak, through a process of legitimation based on the authorization of that violence," acquiring the name "punishment".[87]

This account accurately interprets the text and recognises that punishment is an institution of public law. To this extent, Hobbes is a more modern thinker than his successor, John Locke, whose conception of punishment captures violence both when it is inflicted by the state in accordance with legal prescriptions and when it is inflicted by private persons acting on their

[86] *L*, 14.29, 214/69-70 and 21.14, 338/112. In subchapter 2(a), I explain that Hobbes includes the right to refuse to kill another person within the RSD; this is why a person can no more promise to kill another than to kill himself.
[87] Yates, 247. Throughout his article, Yates methodically lays out the argument summarised in this paragraph.

personal sense of justice.[88] Hobbes's alternative conception does not undermine his overarching ideological objectives, as many scholars have assumed, but reinforces them significantly. The sovereign is the only person in society who retains his entire natural right to inflict violence, and yet the process of authorisation constrains him to channel that right through forms and procedures established by law. These stringent conditions must be satisfied whenever violence is implemented "by publique Authority". In criminal matters, then, the sovereign always acts by authority over juridical subjects. This is why subjects accused of breaking the law have standing to challenge those accusations in a public court of law, quite unlike enemies and rebels.[89] It is also why subjects can never accuse the sovereign of committing an injustice when he punishes them lawfully; after all, they bear final responsibility for his actions.[90]

This careful account of the foundation of punishment is certainly meant to encourage political obedience. It reassures subjects that the treatment they can expect at the hands of the sovereign will be regulated by principles of publicity and legality which are absent in the free-of-all of the state of nature; at the same time, it reprimands those who would complain about the injustice of their punishment by reminding them of their ultimate authorship. And yet while Hobbes does not allow subjects to complain about their punishment, he does allow them to resist it with violence! The question I next examine is whether the execution of punishment in Hobbes's system betrays the promise of legality and obedience contained in its origin story.

---

[88] When describing the state of nature, Locke writes: "the *Execution* of the Law of Nature is in that State, put into every Mans hands, whereby every one has a right to punish the transgressors of that Law to such a Degree, as may hinder its violation…For these two [Reparation and Restraint] are the only reasons, why one Man may lawfully do harm to another, which is what we call *punishment*." (*The Second Treatise of Government*, paragraphs 7-8.)
[89] *L*, 21.19, 342/113 and 28.5, 484/162.
[90] *L*, 18.6, 270/90.

2) The Execution of Punishment

Hobbes's contention that "[the] end of Obedience is Protection"[91] has shocked readers at different times for different reasons. The idea that government deserves our allegiance as long as it guarantees our safety rings scandalously authoritarian in our liberal democratic age. However, we must remember that it is Hobbes's denial of an absolute duty of obedience which outraged many of his contemporaries. Hobbes asserts repeatedly that every human being is entitled to act on his private judgement whenever the sovereign fails to protect him, whether unintentionally (as when the sovereign lacks the capacity to suppress a threat) or intentionally (as when the sovereign himself threatens violence, such as punishment). The subversive implications of this doctrine famously prompted Bishop John Bramhall, a committed royalist, to exclaim: "Why should we not change the Name of *Leviathan* into *Rebells catechism*?"[92]

The uncertain scope of the Hobbesian RSD accounts for this variable reception. At first impression, it seems to be a narrow right with minimal political significance, allowing the individual to resist only serious violence directed against his person. And yet recent scholarship grants this right a much broader scope which would vindicate Bramhall's worry. I begin this chapter by reviewing this scholarship and staking a middle position. Afterwards, I demonstrate that even this moderate interpretation of the RSD creates significant problems for Hobbes's political theory: since it justifies resistance against most forms of punishment, it weakens the sovereign's ability to maintain social order.

[91] *L*, 21.21, 344/114.
[92] Bramhall, 145; cited in Sreedhar, 159. Sreedhar incorporates this quip into the title of her fourth chapter.

I then demonstrate that these practical issues unleash conceptual shockwaves which run all the way down to Hobbes's foundational premises about human rationality. First, I present an original schema of Hobbes's theory of mental illness or, as he puts it, "Madnesse". Second, I show that a subject who breaks the law is mad on this account; but that a subject who thereafter resists punishment, while concurrently authorising that very same punishment, is not mad. I contend that the latter diagnosis is both implausible and inconsistent with various remarks Hobbes makes about madness. To conclude, I reconstruct the RSD in order to mitigate these problems. I argue that Hobbes intends for the right to be opposable to threats of death and extreme suffering only; its extension to punishments which do not necessarily entail these consequences rests on the assumption, eminently reasonable in Hobbes's time, that they would. By shearing the RSD down to its conceptual core, I draw the opposite conclusion to Hobbes: the criminal does *not* have the right to resist most forms of punishment. In this revised theory, the criminal matches the Hobbesian account of madness at every stage of his journey from crime to punishment, and Hobbesian punishment attains harmony in theory and in practice.

a) The Scope of the Right to Self-Defence

Hobbes theorises that we alienate certain rights to the benefit of the sovereign when we enter civil society. It is clear that persons who leave the state of nature give up their right of nature; that is,

> the Liberty each man hath, to use his own power, as he will himselfe, for the preservation of his own Nature; that is to say, of his own Life; and consequently, of doing any thing, which in his own Judgement, and Reason, hee shall conceive to be the aptest means thereunto.[93]

[93] *L*, 14.1, 198/64. Hobbes's conception of natural right connects reason and right. For example, in the state of nature, if reason indicates that keeping a promise will not serve our good, we have the right to break that promise.

It is also clear that these persons do not give up this right in its entirety. In fact, the RSD is best understood as a remainder of the original right of nature which survives in civil society. In this subchapter, I turn to Hobbes's text in order to delineate the precise scope of this right.

Hobbes primarily discusses the RSD in chapter 14, when listing the kinds of rights which persons can never alienate by covenant, and chapter 21, when discussing the liberties retained by subjects. In both instances, his first concern is with life and bodily integrity.[94] As Hampton observes, we can glean a preliminary definition of the RSD as "the privilege or liberty to defend one's body if it is attacked, or to do what is necessary to procure the means (e.g., food and shelter) to assure bodily survival."[95] Sreedhar adds that a subject can exercise this right in the absence of a "direct attack" by a third party; indeed, Hobbes says that one may steal food to avoid starvation and ignore a command to abstain from "food, ayre, medicine, or any other thing, without which he cannot live".[96] The RSD permits a person to do whatever he deems *necessary* to avoid death or bodily injury, in contrast to the full natural right, which also allows him to do whatever he deems *conducive* to that same goal.[97] Many scholars simply assume that this is the proper boundary of the RSD.[98]

However, in the civil state, breaking a promise never serves our good, so we do not have the right to do it: *L*, 14.18-19, 210/68; 15.5, 224/73.

[94] Hobbes begins the three relevant sections as follows: "As first a man cannot lay down the right of resisting them, that assault him by force, to take away his life" (*L*, 14.8, 202/66); "A Covenant not to defend my selfe from force by force, is alwayes voyd" (*L*, 14.29, 214/69); "Covenants, not to defend a mans own body, are voyd" (*L*, 21.11, 336/111).

[95] Hampton, 198-99.

[96] Sreedhar, 8, referring to *L*, 27.25-26, 468/157; see also *L*, 21.12, 336/111-12.

[97] Sreedhar, 15. Necessity connotes a more direct and immediate relationship to the goal than conduciveness.

[98] For example, May, 81-84, only mentions two situations where disobedience is justified, and both feature mortal danger: "laws which require people to serve in battlefield situations, and laws which impose the death penalty..."

Nonetheless, Hobbes very quickly brings other threats within the ambit of the RSD. The most obvious is imprisonment: in chapter 14, Hobbes first puts the risk of death on par with "Wounds and Chayns, and Imprisonment"[99], and then repeats a similar list according to which "no man can transferre, or lay down his Right to save himselfe from Death, Wounds, and Imprisonment".[100] In the same chapter, he affirms that a subject can refuse to incriminate himself *or his loved ones* in a criminal proceeding; the latter claim expands the scope of the right beyond mere preservation of life and limb.[101] In chapter 21, Hobbes enlarges this right further still: for instance, he says that a subject may refuse to kill other people. Reviewing these passages, Hampton discards her preliminary definition: "The self-defense right has now been interpreted so broadly that is it essentially equivalent to the *entire* right to preserve oneself."[102]

While this claim is not true, strictly speaking – Hobbes expects subjects to obey the limits placed by civil law on their natural right in most circumstances – Hampton is obviously correct that the RSD extends beyond physical threats. In fact, scholars have recently made the rather counterintuitive case that Hobbes propounds a broad theory of *resistance* to political authority. Before delving into the particulars, I should make a terminological point. These revisionist scholars generally apply the term "RSD" to the right which meets Hampton's preliminary definition; they craft different terms for what they take to be broader cognate rights in Hobbes's theory. Thus, Eleanor Curran speaks of a "right to full preservation" held in civil society.[103] Sreedhar, who purports to discover a unified theory of resistance in *Leviathan*, devotes one

[99] *L*, 14.8, 202/66.
[100] *L*, 14.29, 214/69-70.
[101] *L*, 14.30, 214/70.
[102] Hampton, 201.
[103] Curran, 105.

chapter to the narrow RSD, another to the "true liberties of subjects" presented in chapter 21, and another to the right of rebellion.[104] However, I use a single term – the RSD – for the subject's entitlement to resist sovereign authority in all instances. At present, I can only justify this choice with an appeal to linguistic simplicity. In subchapter 2(d), I show that the instances which give rise to the right all share a common theme; thus, it is conceptually appropriate to speak of a single right with various manifestations, as opposed to a variety of rights.

Returning to the scope of this right, Hobbes mentions several instances where the subject is entitled to resist authority – either passively (by refusing to obey a command) or actively (by acting to frustrate public officials). These include instances where obedience would entail:

1. Death;[105]
2. Bodily injury;[106]
3. Imprisonment;[107]
4. Bearing witness against oneself in a criminal proceeding;[108]
5. Bearing witness against someone "by whose Condemnation a man falls into misery; [such as] a Father, Wife, or Benefactor" in a criminal proceeding;[109] or
6. Killing another person.[110]

In these instances, the RSD is absolute. Hobbes also claims that the right applies *conditionally* whenever the sovereign commands the subject to execute

7. "[Any] dangerous, or dishonourable Office".[111]

---

[104] These are chapters 1, 2, and 4, respectively.
[105] *L*, 14.8, 202/66; 14.29, 214/69-70; 21.11-12, 336/111-12; 21.14-15, 338/112; 27.25-26, 468/157; 28.2, 482/161.
[106] Most of the passages listed in the previous footnote place bodily injury on par with death.
[107] *L*, 14.8, 202/66; 14.29, 214/69-70.
[108] *L*, 14.30, 214/70 (this includes the right to lie under torture); 21.13, 337/112.
[109] *L*, 14.30, 214/70.
[110] *L*, 21.14-15, 338/112.

In such a case, "When...our refusal to obey, frustrates the End for which Soveraignty was ordained; then there is no Liberty to refuse: otherwise there is."[112] Hobbes's much discussed example of the military draft helps illuminate his meaning. He says that a person is entitled to dodge the draft if he can find a substitute, "for in this case he deserteth not the service of the Common-wealth", even if his decision rests on a base motive such as "Cowardise". But if the commonwealth faces grave danger which requires the entire population to bear arms, then no subject can refuse to do so, "because otherwise the Institution of the Common-wealth, which [the subjects] have not the purpose, or courage to preserve, was in vain."[113] In other words, the subject may refuse to expose himself to danger or dishonour *except* when refusal imperils the institution which habitually guarantees his safety. The assumption of risk is only justified by the prospect of much greater risk.

Sreedhar concludes that this nexus of entitlements establishes "a general right to resist the punishment commands of the sovereign."[114] In a footnote, she clarifies that it only applies to three categories of punishment: capital punishment, corporal punishment, and imprisonment. She reasons that there is no right to resist pecuniary penalties "unless, of course, one can reasonably see payment of those fines as posing a threat to one's self-preservation."[115] Nevertheless, the RSD is clearly opposable to *most* forms of punishment – in fact, it is opposable to the *most serious* forms, which presumably attach to the most serious crimes. One might suppose that this interpretation of the RSD represents a major obstacle to the execution of punishment and, by

[111] *L*, 21.15, 338/112. When discussing the "true liberties of subjects", Sreedhar, 53-88, distinguishes "unconditional" from "conditional" liberties. The former correspond to points 1-6 on my list, the latter to point 7.
[112] *L*, 21.15, 338/112.
[113] *L*, 21.16, 338-40/112.
[114] Sreedhar, 73.
[115] *Ibid*. at footnote 40.

extension, the preservation of civil society. However, scholars typically shrug off this concern on the basis that individual attempts to evade punishment cannot impair a well-ordered commonwealth; the idea is that the sovereign should have the material capacity to bring criminals to justice without their cooperation or the cooperation of their loved ones.[116] As Deborah Baumgold puts it, the RSD is "politically irrelevant" and "inconsequential" because it does not allow for organised resistance.[117]

Or does it? Glenn Burgess claims that Hobbes advances a theory of resistance with two dimensions. The first is the individualistic dimension just outlined. The second, however, has political implications: it allows the subject to join forces with others and declare a rebellion whenever the sovereign breaches the laws of nature so egregiously that he positions himself in a state of war against his subjects.[118] Sreedhar agrees and attempts to delineate a Hobbesian right of rebellion in the final chapter of her book.[119] She correctly notes that any such right must be understood as an *individual right to collective action* as opposed to a right held by a collective, since the latter would be inconsistent with Hobbes's legal theory.[120] Sreedhar articulates this purported right of rebellion as follows: "a subject has the right to rebel if and only if that subject judges that the sovereign is not providing adequately for his security and that rebellion is the best means for self-preservation."[121] She specifies that a subject may deem his life insecure if he (1) reasonably fears violence at the hands of other persons, whether private individuals or

[116] See *e.g. ibid*., 72 and 77-78.
[117] Baumgold, 31-35.
[118] See generally Burgess.
[119] Sreedhar, 132-67.
[120] *Ibid.*, 148-50. As explained in footnote 57, Hobbes denies personality (and the capacity to hold and exercise rights) to a "Multitude of men" – that is, to a group without a designated representative: *L*, 16.13-14, 248-50/82.
[121] Sreedhar, 137.

government agents, or (2) lacks one or more of the basic necessities of life. Sreedhar acknowledges the various arguments Hobbes makes against rebellion but claims that they only prohibit uprisings motivated by religious or political ideology, as opposed to what she calls "rebellions from necessity".[122]

These arguments fail for two reasons. First, a right of rebellion is unnecessary to accommodate situations where the sovereign intentionally and unlawfully threatens the security of his erstwhile subjects, such as genocides. As Burgess suggests, a sovereign who engages in such behaviour abdicates his official duties towards these individuals and restores them to the state of nature.[123] This is different from punishment, even in its capital variety: as demonstrated in chapter 1, punishment inflicted according to law is grounded in the authorisation of every subject, even the one who is targeted. By contrast, the arbitrary infliction of force against an individual treats him like an enemy in the state of nature.[124] But in the state of nature, the individual need not appeal to a residual right of rebellion; he recovers his entire natural right and can justly resist his former sovereign through any action he deems necessary. Whether he creates a wartime alliance with others in furtherance of this goal has no conceptual import.

Second, the distinction Sreedhar proposes between ideology and necessity is not firm enough to limit the right of rebellion in a manner consistent with Hobbes's commitment to

[122] *Ibid.*, 142-43.
[123] At the end of chapter 21, Hobbes lists various situations where the subject is released from his political bonds because he can no longer depend on the sovereign for protection. Formal abdication is among them: *L*, 21.23, 344/114. In my view, a sovereign who attacks his subjects abdicates his office *de facto*, since he repudiates the overarching function of sovereignty – protection.
[124] Hobbes draws a firm distinction between the punishment of subjects, which proceeds from authorisation, and the suppression of rebels, which proceeds from natural right. See especially *L*, 28.13, 486/163 and 28.22-23, 492-94/165-66, discussed in footnote 77 and the corresponding text. Extra-judicial persecution appears juridically analogous to the suppression of rebels. Indeed, the only difference is factual: in the first case, responsibility for breaking the political bond lies with the sovereign; in the second, with the rebels.

political stability. It is evident that the RSD permits a subject to fight off "the terrour of present death" in civil society,[125] for example by stealing medicine or knocking out a knife-wielding assailant. However, these instances fall within the scope of the RSD as previously defined. It is conceptually irrelevant that the endangered subject might sometimes exercise this right in concert with other similarly endangered subjects.[126] Members of a group who exercise their RSD against an immediate threat at the same time do not become a rebellious faction, as they do not thereby repudiate their political obligations.

Of course, these are not the situations Sreedhar has in mind. Rather, she seems concerned that the sovereign might be chronically unable to guarantee his subjects' security. But even here, it is unnecessary to speak of a retained right of rebellion. Hobbes concedes that a commonwealth is always in danger of breaking down. However, when the sovereign loses the ability to maintain social order, the political obligations of subjects dissolve and they return to the state of nature.[127] Should the former sovereign attempt to impose his authority upon them anew, they would be entitled to resist pursuant to their right of nature. The problem with Sreedhar's argument is the further implication that subjects who remain in civil society might hold a residual right to rebel if they judge that their sovereign's *general policies* do not adequately provide for their security.[128] This is distinct from the right to repel immediate threats, which every subject obviously holds.

---

[125] *L*, 27.25, 468/157.

[126] For instance, he might join forces with his friends in a bar to hold off a group of aggressive strangers until the police arrive.

[127] Hobbes discusses this eventuality throughout chapter 29 of *Leviathan*, 498-519/167-74, entitled: "*Of those things that Weaken, or tend to the* DISSOLUTION *of a Common-wealth*". See also *L*, 21.21-25, 344-46/114-15 for instances where the sovereign's inability to protect extinguishes the subjects' duty to obey.

[128] Sreedhar's definition of the right, 137, appears to make room for this possibility: "a subject has the right to rebel if and only if that subject judges that the sovereign is not providing adequately for his security and that rebellion is the best means for self-preservation."

But what of subjects who believe their sovereign's foreign policy is too aggressive and threatens them with the prospect of war? What of those who disagree with legal restrictions on firearm ownership, on the basis that these measures hinder them from protecting their homes from violent criminals?[129] The distinction between ideology and necessity falls apart in the absence of immediate threats. And while the RSD clearly allows subjects to resist such threats, it is far from clear that it also allows them to rise up against their sovereign to pre-empt future threats which his political judgement might produce.

Ultimately, Sreedhar's defence of the right to rebel assumes that subjects may resist the sovereign if they are dissatisfied with their quality of life. In subchapter 2(d), I show that even her understanding of the traditional RSD rests on this assumption, which is mistaken. Nevertheless, even if the RSD is less capacious than Sreedhar believes, it indisputably permits the subject to resist most forms of punishment. This includes the standard punishment for serious offences in contemporary Western states: imprisonment. Moreover, this right might not be as harmless as most scholars believe.[130] Quite obviously, it can only heighten the incidence of violence in society. Hobbes intends for his theory to be pedagogical; by advising subjects to resist most kinds of punishment, he increases the likelihood that violent crimes will be followed by violent resistance which might claim additional victims (passers-by, police officers, the criminal himself).

[129] Of course, other subjects might reach the opposite conclusion: that *failure* to restrict firearm ownership endangers them by increasing the number of weapons in circulation. Sreedhar's right of rebellion confronts the sovereign with the prospect of justified rebellion by *some* subjects irrespective of his approach to this controversial policy issue – and many others like it. In subchapter 2(d), I argue that Hobbes does not advocate this right because he is committed to subordinating the political judgement of individuals to a common authority.

[130] Recall Baumgold's claim that the RSD is "politically irrelevant" and "inconsequential".

Furthermore, even this modest version of the RSD challenges Hobbes's understanding of human rationality in two important ways. First, Hobbes presents the social contract as the most rational solution to the problem of interpersonal violence. However, a prospective subject might calculate that there is a rather high probability he will eventually break some law and become eligible for punishment. If he shared Hobbes's views on self-defence, he would foresee that he would afterwards square off against a mighty sovereign in a violent struggle. This prospect would likely appear less attractive to a rational agent than continued exposure to danger from disorganised rivals in the state of nature. On this understanding of the RSD, the social contract can be defended either by appeal to non-rational considerations or not at all.

The second challenge to Hobbesian rationality runs even deeper. I recall that Hobbes's theory of authorisation holds that the "words and actions" of the sovereign can be attributed to every one of his subjects.[131] Since a criminal has authorised his own punishment, he thereby "owns" the sovereign's will to punish him. At the same time, Hobbes implies that the criminal should form a personal will to resist any punitive action taken against him. On Hobbes's theory of personality, then, the criminal owns *two contradictory wills* at the same time. Serious questions arise about the rationality of this individual, especially in light of two claims made by Hobbes: that it is absurd to contradict oneself and that absurdity "may also be numbred amongst the sorts of Madnesse".[132] To tackle these issues, I first tease out a theory of madness from *Leviathan* and then examine the two-willed criminal through its lens.

[131] *L*, 16.4, 244/81. Hobbes's theory of authorisation is canvassed in subchapter 1(a).
[132] *L*, 14.7, 200-02/65 and 8.27, 122/39.

b) A Hobbesian Theory of Madness

While Hobbes's professed rationalism has caused numerous scholars to study his conception of reason, its antithesis – which he calls "Madnesse" – remains surprisingly undertheorised. Indeed, there is no treatment of Hobbesian madness which is both comprehensive and systematic in the academic literature;[133] this subchapter is intended as a modest first step towards filling this gap. My primary claim is that Hobbes conceives of madness as the faulty operation of reason within deliberation. I contend that his oft-ignored remarks on madness provide evidence valuable to two central debates about his theory of practical deliberation. In this subchapter, I delineate a satisfactory interpretation of that theory; in the next, I rely on it to evaluate the Hobbesian criminal's rationality.

The bulk of Hobbes's remarks on madness in *Leviathan* are found in chapter 8. He devotes the first half of the chapter to those "abilityes of the mind, as men praise, value, and desire should be in themselves".[134] These are "naturall wit", consisting in the ability to form thoughts quickly and to give them a "*steddy direction* to some approved end",[135] and "acquired wit", consisting in the proper exercise of scientific reason.[136] After asserting that people differ in wit because they feel different passions,[137] Hobbes transitions to a discourse on madness.

---

[133] Numerous scholars have produced excellent work focusing on specific aspects of Hobbesian madness, and I have referenced several (but not all!) of them in this subchapter. Others take a broader view but fail to provide a systematic account: see *e.g.* the generally accurate textual summary in Weber. Hampton, 34-42, comes closest to offering a comprehensive explanatory theory; unfortunately, she seriously misinterprets Hobbes, as I show in the text corresponding to footnotes 190-98.
[134] *L*, 8.1, 104/32.
[135] *L*, 8.2, 104/32.
[136] *L*, 8.13, 110/35.
[137] *L*, 8.14-15, 110/35.

First, Hobbes situates madness at one end of the spectrum charting the intensity of passionate feeling.

> And therefore, a man who has no great Passion for any of these things [sorts of power], but is as men terme it indifferent; though he may be so farre a good man, as to be free from giving offence; yet he cannot possibly have either a great Fancy, or much Judgement. For the Thoughts, are to the Desires, as Scouts, and Spies, to range abroad, and find the way to the things Desired: All Stedinesse of the minds motion, and all quicknesse of the same, proceeding from thence. For as to have no Desire, is to be Dead: so to have weak Passions, is Dulnesse; and to have Passions indifferently for every thing, GIDDINESSE, and *Distraction*; and to have stronger, and more vehement Passions for any thing, than is ordinarily seen in others, is that which men call MADNESSE.[138]

He proclaims a reciprocal relationship between our abnormal passions and our sensory organs.[139] He then identifies two passions that are particularly conducive to madness: "vaine-glory" (also called "pride" or "selfe-conceipt"), which causes a madness called "rage" or "fury"; and "dejection", which causes a madness called "melancholy".[140]

Hobbes makes clear that a passion can generate madness even if not felt in excess, provided that it points to an evil outcome.

> In summe, all Passions that produce strange and unusuall behaviour, are called by the generall name of Madnesse. But of the several kinds of Madnesse, he that would take the paines, might enrowle a legion. And if the Excesses be madnesse, there is no doubt but the Passions themselves, <u>when they tend to Evill</u>, are degrees of the same.[141]

From the latter prong of this definition, he affirms that rebels are afflicted with the madness of rage, "For they will clamour, fight against, and destroy those, by whom all their life-time before, they have been protected, and secured from injury."[142] The problem here is that "the Passions

---

[138] *L*, 8.16, 110/35.
[139] *L*, 8.17, 112/35. This poses no difficulty because Hobbes thinks that our mental states can be reduced to bodily movements: see *e.g. L*, 1.1-4, 22-24/3-4 and *L*, 6.1, 78/23. Therefore, an abnormal organ can produce movements which translate into abnormal thoughts, just as abnormal thoughts are reducible to movements capable of harming the physiology of an otherwise healthy organ.
[140] *L*, 8.18-20, 112/35-36.
[141] *L*, 8.20, 112/36 (emphasis mine).
[142] *L*, 8.21, 112/36.

themselves", even if not excessively vehement, have an *abnormal content*; if pursued, they produce "Evill" outcomes for the agent.

Hobbes further states that the "passions unguided" are a form of madness, offering the drunkard and the insouciant daydreamer as examples.[143] He then defends his theory linking madness to the passions against an apparently popular rival that posited a link to demonic possession.[144] Hobbes concludes with a frontal assault against his scholastic opponents, asserting that the absurdities committed by those who abuse words "may also be numbred amongst the sorts of Madnesse."[145] This coheres with his suggestion, made in passing earlier, that unguided thoughts are a form of madness.[146]

These passages show that Hobbes regards as mad at least four categories of persons:

1. those with passions of *abnormal intensity*;
2. those with passions with *abnormal content*;
3. those with *unguided passions*; and
4. those with *unguided or absurd thoughts*.

These categories cannot be understood outside Hobbes's general theory of practical deliberation. How do thoughts and passions influence the actions of sane persons? What role, if any, does reason play in their deliberation? Quite obviously, we must grasp his views on these issues to see why he calls certain people mad. And yet this is no easy task – the complexity of Hobbes's thought has raised serious disagreement on this topic among his interpreters. It is not

[143] *L*, 8.23, 114/36-37.
[144] *L*, 8.24-26, 114-120/37-39.
[145] *L*, 8.27, 122/39.
[146] *L*, 8.3, 106/33.

my objective to exhaustively reconstruct Hobbes's theory of action. Rather, I will review two existing debates in the scholarly literature and ask whether Hobbes's remarks on madness help to resolve them. Although I do not claim that these remarks are decisive, I suggest that they support a model that is independently superior to its competitors. Its agreement with Hobbes's theory of madness stands as one more argument in its favour.

First, a terminological point. Hobbes appears to use the word "reason" in various ways and so, for the sake of clarity, I will adopt Bernard Gert's threefold taxonomy. "Instrumental reason" refers to the ability to discover means to attain whatever end we happen to hold, irrespective of its content.[147] "Verbal reason" refers to the ability to logically combine linguistic propositions, once again without any concern about the end we hope to achieve.[148] "Natural reason" refers to a faculty which attracts us to certain particular ends: on Gert's uncontroversial reading of Hobbes, these are the avoidance of pain and death.[149]

The first contested question in the literature is whether reason can motivate us to pursue certain ends. Those who interpret Hobbesian reason as purely instrumental or verbal (or both) deny it, whereas those who make room for natural reason do not. I refer to the first position as the *passive* account because it holds that reason stays silent during will-formation: the passions set our goals and we do not reflect upon them. I refer to the second position as the *active* account because it holds that reason identifies certain passions as important and somehow causes us to

[147] Gert, 244.
[148] *Ibid.*, 245-46. This is the definition Hobbes gives when he formally defines reason: *L*, 5.2, 64/18.
[149] Gert, 248.

give them priority, even in the face of contrary passions. Importantly, the active account does not dispute that reason can also play an instrumental or verbal role once we have chosen our ends.

The second question concerns how reason is positioned towards deliberation. Some writers argue that these processes operate separately: I refer to this position as the *exclusivist* account. Others argue that reason interacts with the passions during deliberation: I refer to this position as the *participative* account.

I will put my cards on the table right away. In my view, Hobbes's remarks on madness presuppose an *active* and *participative* role for reason in deliberation. In other words, natural reason enters the deliberative process and leads us to respond to our self-preservation as a reason for action. At the outset, it is useful to look at what Hobbes says when he introduces the concepts of deliberation and the will in chapter 6. He begins with a general definition of deliberation:

> When in the mind of man, Appetites, and Aversions, Hopes, and Feares, concerning one and the same thing, arise alternately; and divers good 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.[150]

He also offers the following clarification:

> And because in Deliberation, the Appetites, and Aversions are raised by foresight of the good and evill consequences, and sequels of the action whereof we Deliberate; the good or evill effect thereof dependeth on the foresight of a long chain of consequences, of which very seldome any man is able to see to the end.[151]

[150] *L*, 6.49, 90/28.
[151] *L*, 6.57, 94/29.

Hobbes stresses that deliberation only ends when the deliberator performs the action deliberated upon or when performance becomes impossible.[152] He then defines the will as follows:

> In *Deliberation*, the last Appetite, or Aversion, immediately adhaering to the action, or to the omission thereof, is that we call the WILL; the Act, (not the faculty,) of *Willing*. [...] *Will* therefore *is the last Appetite in Deliberating*.[153]

Hampton (who offers the most comprehensive discussion of Hobbes's theory of madness)[154] reads these passages in favour of the passive and participative accounts. She accepts – as do I – that reason plays a part in deliberation, but she views its function as merely instrumental. She argues that the deliberator's mental debate "appears to be between or among desires *alone*," and that "reason's only role in the deliberative process is to help determine how to achieve a goal set by desire – it does not itself dictate a goal nor motivate us to pursue it."[155]

In the passive camp, Hampton finds herself in good company. Upon parsing the primary text, John Deigh defends a strictly verbal conception of reason.[156] Thomas Pink reaches a similar conclusion, although he interprets Hobbes against the background of his intellectual context, particularly his disagreement with scholastic theories of the will.[157] Stephen Darwall advances an instrumental view of reason that gives it no more motivational force.[158] Terence Irwin, Patrick Riley, and Samantha Frost also seem to adhere to the passive view.[159] However, those who

---

[152] *L*, 6.52, 92/28.
[153] *L*, 6.53, 92/28.
[154] See Hampton, 34-42, for the discussion, and the text corresponding to footnotes 190-98 for critique.
[155] *Ibid.*, 19. See also 35: "reason aids a deliberator in determining causal connections, but it does so in the service of that person's passions, which alone can move him to act."
[156] See Deigh. Hoekstra (2003) convincingly rebuts this view.
[157] See generally Pink (2003, 2011a, 2011b, 2016).
[158] Darwall (1995), 59: "[Reason] can recommend no conduct or end directly or intrinsically. Its practical function is purely instrumental, to work out the means or 'way to the thing desired.'"
[159] Irwin, 105; Riley, 43-44; Frost, 101.

defend the passive account typically disagree with Hampton that reason acts within deliberation. They are exclusivists who view deliberation as a succession of purely conative states, one of which overcomes the others due to its intensity and becomes the will. Examples of this stance can be found in Pink, Darwall, and Irwin.[160]

Adrian Blau provides an interesting segue to the active account because he endorses it in addition to exclusivism. More specifically, he thinks that reason helps determine our ends *before* we deliberate.[161] Although he rejects Gert's definition of natural reason,[162] Blau asserts that self-preservation is our "real good" due to our biological makeup. Nevertheless, people sometimes desire "apparent goods" which undermine their real good.[163] The role of reason, then, is to help passions that advance our real good win the day against passions for dangerous apparent goods.[164] Blau rightly points out that reason can alter the imagination by showing us the empirical or logical consequences of our actions, or by modifying our underlying beliefs. These thoughts influence the passions we experience in deliberation and help determine which passion

[160] Pink (2016), 172 ("Hobbes denied the existence of distinctively intellectual and action-constitutive motivations of the will...") and 185 ("There are no longer distinctively reason-involving motivations. All motivations are passions...") and generally Pink (2011b); Darwall (2000), 331, footnote 3 ("Hobbes makes no place for critical reflection in his account of deliberation..."); Irwin, 105 ("This account of deliberation refers only to non-normative states...").
[161] Blau, 209-12. Although Blau's characterization of reason as "deductive" matches the verbal definition, 196-97, he insists that reason can influence our imagination and change the deliberator's will, 205-06, 211. That is why I consider him a proponent of the participative position.
[162] *Ibid*., 207-08.
[163] For example, eating a cake might appear good to me, but if it is going to dangerously increase my blood sugar level, it subverts my "real good". The language of "apparent good" is from *L*, 6.57, 94/29.
[164] Blau, 212.

ends up as our will.[165] Reason is thus a "counselor" to the passions, predisposing us to favour those which secure our self-preservation.[166]

Those who remain are the defenders of the active and participative combination. I start with the *participative* position because it is easier to establish its superiority to its rival. Considered participative theories begin with Hobbes's mention of "divers good and evill consequences"[167] and "foresight of the good and evill consequences, and sequels"[168] of an action when he defines deliberation. They also take into account his statement in chapter 8 that "the Thoughts, are to the Desires, as Scouts, and Spies, to range abroad, and find the way to the things Desired".[169] They read these passages as evidence that cognitive processes occur within deliberation, and turn to the discussion of "mentall discourse" (also called the "trayne of thoughts") in chapter 3 to understand their character. That chapter distinguishes mental discourse which is unguided[170] from that which is "*regulated* by some desire, and designe."[171] It then divides the latter category into two branches: the seeking of causes which can produce an effect, and the seeking of the possible effects of a thing or action.[172] The first amounts to instrumental reason and the second to prudence (if we rely only on our experience) or verbal reason (if we rely on scientific theorems).[173] On the participative account, Hobbes's comments in chapters 3

[165] *Ibid.*, 209-12, referring among other passages to *L*, 6.57, 94/29.
[166] Blau, 214-16.
[167] *L*, 6.49, 90/28.
[168] *L*, 6.57, 94/29.
[169] *L*, 8.16, 110/35.
[170] *L*, 3.3, 38/9.
[171] *L*, 3.4, 40/9.
[172] *L*, 3.5, 40-42/9-10.
[173] See *e.g. L*, 3.7, 42-43/10; 5.21, 76/22; 8.11, 108/34.

and 8 incorporate instrumental reason, verbal reason (for those who have attained it), and prudence (for those who have not) *into* deliberation.

In this vein, Kinch Hoekstra asserts that, "If reason is understood as a mental capacity of humans, as Hobbes repeatedly understands it, it is only an improved form of the train of thoughts."[174] Laurens van Apeldoorn agrees that deliberation implies this idea of reason.[175] He insightfully remarks that if mental discourse can be regulated by "some desire", then it must be analogous to deliberation, since Hobbes gives no indication that we ever feel any desires outside deliberation. Rather, "it is more probable that he took all mental discourse in which we consider the consequences of our actions and that produce appetites and aversions in us as instances of deliberation."[176]

Abizadeh also insists that Hobbesian deliberation involves cognitive processes, which include reasoning among humans.[177] He describes the passions as "hybrid mental states" that include (1) cognitive representations of a given object *as* a good or evil object and (2) a conative urge to favour or disfavour that object.[178] The deliberator experiences passions formed by cognitive representations which he evaluates before acting. In my view, the textual evidence marshaled by these authors is decisive. Tellingly, Blau can only defend exclusivism through a

[174] Hoekstra (2003), 117.
[175] van Apeldoorn (2012), 146.
[176] *Ibid.*, 154.
[177] Abizadeh (2017), 3.
[178] *Ibid.*, 11.

strained reading of chapters 3 and 8 of *Leviathan*, which he argues have nothing to do with Hobbes's conception of reason.[179]

The participative account does not necessarily entail the *active* account, as it is possible to argue that deliberation incorporates passive reasoning.[180] Hampton's purely instrumental view of reason fits the bill: this faculty purportedly enters deliberation only to show us how we can achieve the ends proposed by our various desires, without evaluating their normative worth.[181] However, significant textual evidence gainsays this reading. Recall Hobbes's statement that "the Appetites, and Aversions are raised by foresight of the good and evill consequences, and sequels of the action whereof we Deliberate",[182] which implies that reason can affect our desires by showing us the likely outcomes of potential behaviour. This ties into his broader claim that "Voluntary motions, depend alwayes upon a precedent thought",[183] which establishes a causal relationship from thought to action. At another point, Hobbes flatly asserts that "the Actions of men proceed from their Opinions".[184] And what are "opinions"? They are the thoughts that succeed one another in mental discourse, culminating in a "judgement".[185] If mental discourse

[179] Blau, 201-04. I cannot critique Blau's reading here. Suffice it to say that his arguments sit uneasily with the textual passages cited by Abizadeh, Hoekstra, and van Apeldoorn.
[180] Abizadeh (2017), 13, restates this argument (with which he disagrees) in these terms: such reasoning would be experienced as "a purely passive, albeit cognitive, process, i.e. that it involves a succession of mental states one simply experiences following innate associative structures, and not mental states that are somehow reflectively endorsed."
[181] Hampton, 19.
[182] *L*, 6.57, 94/29 (emphasis mine).
[183] *L*, 6.1, 78/23.
[184] *L*, 18.9, 272/91.
[185] *L*, 7.2, 98/30.

plays a part in deliberation (as the participative model claims), then it must be an active part, helping constitute our passions by judging their effects as good or evil.[186]

But are these judgements arbitrary, or does reason inevitably lead us to value certain outcomes over others? For Gert, natural reason orients our passions towards our self-preservation so that we only deliberate rationally if we pursue that end. Drawing chiefly on passages from *De Cive*, he concludes that "Hobbes would regard someone who uses all of his experience, instrumental reasoning, verbal reasoning, and science in order to kill himself in the most painful possible way, not only as mad, but as acting irrationally."[187] Hoekstra agrees that "Hobbes does think of reason itself as natural, or dependent on a directive desire": namely, self-preservation.[188] This is why Hobbes claims that our "right of nature" consists in a liberty to do anything "which, in [our] own Judgement, and Reason," will protect our lives; why he crafts a long list of "laws of nature", which are "found out by Reason", whose purpose is to promote our self-preservation; and why he tells the infamous fool that no action that "tendeth to [our] own destruction" can be "reasonably or wisely done."[189]

Hobbes's comments on mental illness clearly presuppose the active account of reason. When he says that madmen experience passions with an abnormal intensity or content, he can be understood as suggesting that self-destructive passions somehow override natural reason within their minds. But if reason did not evaluate self-destructive passions as undesirable among the

[186] Abizadeh (2017), 4-5, and van Apeldoorn (2012), 164, hit on these passages when defending the active account. van Apeldoorn (2014), 617, explains that Hobbes distinguishes these consequence-sensitive motivations from instinctive urges, despite rejecting the scholastic term "rational appetites" at *L*, 6.53, 92/28.
[187] Gert, 248.
[188] Hoekstra (2003), 120.
[189] *L*, 14.1, 198/64; 14.3, 198/64; 15.5, 224/73. Hoekstra (2003), 119, cites these passages and many others to make this point.

sane, what sense would it make to call people who act on those passions insane? Nonetheless, Hampton attempts to reconcile Hobbes's comments on madness with the passive account. She advances three arguments. First, she says that Hobbes espouses "true belief instrumentalism", and so he condemns as mad people who form a proper desire but act against it due to mistaken beliefs.[190] This is false: Hobbes would say that these people simply committed an error, something "to which even the most prudent men are subject",[191] without questioning their sanity. Second, Hampton claims that some madmen properly form a desire for self-preservation but then experience "certain extreme bodily motions" which provoke contrary passions that "usurp the relational pursuit of [their] predominant, ruling passion for [their] own preservation."[192] This is incoherent: it is difficult to understand how one of the various passions in deliberation can be considered "predominant" if it does not end up as the deliberator's action-producing will.

Finally, Hampton alleges that even Hobbes's content-based criticism of certain passions agrees with instrumental reason, since those problematic passions are "produced by a diseased and abnormal physiological state, such that [madmen] cannot rationally pursue the object they *naturally* want when they are not in this diseased state."[193] As such, madness results from "massive biological misfiring"[194] which produces "wrong" or "spurious" desires;[195] it cannot be attributed to some failure of reason to channel our desires. But Hampton's insistence that reason cannot influence our desires because they have a physiological basis implies that reason can be separated from physiology. Such a claim is unwarranted: as a thoroughgoing materialist, Hobbes

[190] Hampton, 36-37.
[191] *L*, 5.5, 68/19.
[192] Hampton, 38.
[193] *Ibid.*, 39.
[194] *Ibid.*, 40.
[195] *Ibid.*, 41.

attributes a physiological basis to *all* our mental phenomena.[196] The active account of reason does not suppose that some immaterial substance guides our passions.[197] The issue of whether rational cognitive processes motivate our actions can be bracketed from the issue of whether a theory of action is materialistic.[198]

At long last, it is time to interpret Hobbes's account of madness through the lens of active, participative reason. Some kinds of madness arise from reason's inability to fulfil its verbal or instrumental roles. These are the afflictions of those with unregulated thoughts or who commit absurdities: they cannot order their thoughts into logical sequences and so they deliberate unsoundly.[199] The predicament of those with unguided passions is similar, as Hobbes draws no firm distinction between mental discourse and deliberation.[200] When he connects the drunkard's and daydreaming wanderer's aimless passions to madness, without addressing their intensity or content, Hobbes should be understood as castigating their failure to reason about consequences. As a result, these persons' actions result from random instincts that they never subjected to cognitive evaluation. Hobbes implies that such evaluation is essential to sanity.

What of the passions that are "stronger, and more vehement" than ordinary or that "tend to Evill"?[201] These categories can be collapsed: the underlying point is that some passions motivate us to harm ourselves. The first category contains those passions that are harmful only if

[196] *L*, 1.1-4, 22-24/3-4; 6.1, 78/23.
[197] Hobbes thinks this very term is absurd: *L*, 5.5, 68/19.
[198] van Apeldoorn (2012), 146.
[199] According to the active-participative model, reason shows us the various consequences of an imagined action, and each of these representations triggers certain passions. Hobbes is saying that there should be coherence between the subjective cognitive evaluation of a given situation, the imperative for self-preservation, and the action-producing passion within a healthy mind.
[200] Recall that mental discourse, like deliberation, is regulated by "some desire, and designe": *L*, 3.4, 40/9. See also van Apeldoorn (2012), 154.
[201] *L*, 8.16, 110/35; 8.20, 112/36.

felt in excess[202] and the second those that are always harmful.[203] For Hobbes, such passions cannot constitute the will of a sane deliberator. In a healthy mind, reason does not only illustrate the consequences of our passions; it also draws us towards those that coincide with our self-preservation.[204]

In sum, Hobbes's theory of madness best coheres with an active, participative role for reason in deliberation. Reason should (1) represent to us the likely consequences of each passion we feel, (2) order those representations logically, and (3) help us respond to our self-preservation as a reason for action. In the mind of a madman, this faculty fails to fulfil at least one of these tasks. What of the mind of a criminal?

c) Diagnosing the Lawbreaker and the Fugitive

The word "schizophrenia" is derived from the Greek expressions σχίζειν ("to split") and φρήν ("mind"). This etymology accounts for the popular misconception that a schizophrenic is a person whose mind is split between different personalities. However, the clinical definition of schizophrenia describes a condition where the mechanism linking feeling, thought, and action is disturbed within a *single* personality. Those who experience a mind split between *different*

[202] For instance, hunger motivates eating (which promotes our self-preservation) but excessive hunger motivates overeating (which is harmful).
[203] According to Hobbes, vain-glory is always detrimental, as it has no discernible benefit and often provokes interpersonal conflict: see *L*, 8.18-19, 112/35-36; 27.13, 460/154; 27.17, 462/154-155. Vain-glory is not a more intense variant of glory. The difference is that glory is grounded on a correct estimation of one's abilities and vain-glory on an erroneous one: *L*, 6.39, 88/26-27.
[204] Again, this is what Gert calls "natural reason". Of course, several competing passions might all be consistent with our self-preservation. Once it has filtered out our noxious passions, natural reason does not appear to play any further role. The intensity of each remaining passion will determine its ranking in our deliberation.

personalities are said to suffer from "multiple personality disorder" (or, to use a more recent term, "dissociative identity disorder").[205]

I am not bringing up this scientific knowledge in order to apply contemporary diagnostic criteria to Hobbes's criminal. Indeed, whether a present-day psychiatrist would consider such a patient to be schizophrenic or dissociated (or both) is of little hermeneutic value. My purposes are purely illustrative. By keeping these categories in mind when examining the criminal in light of Hobbes's own theory of madness, we will more readily understand the mental states his theory captures and excludes. We will then be better able to probe the plausibility and internal coherence of that theory.

Hobbes clearly considers the criminal mad when he breaks the law,[206] as he thereby undermines his short-term or long-term self-interest.[207] The former occurs when the criminal does not believe that he will escape punishment but breaks the law anyway. On Hobbes's definition, laws are paired with punishments that inflict greater harm than the expected benefits of transgression.[208] If the criminal produces coherent thoughts which point to this outcome but fails to respond to his own well-being as a reason for action, he fits Hobbes's definition of a

[205] Statt, "schizophrenia", "multiple personality"; Noll, "multiple personality and schizophrenia".
[206] Hobbes explicitly condemns rebels as mad for breaching their political obligations: *L*, 8.21, 112/36. These comments apply *mutatis mutandis* to non-political criminals. Interestingly, Hobbes also claims that "Mad-men" do not bear responsibility for their actions, and therefore that they stand outside the scope of the civil laws and punishment: *L*, 16.10, 248/82; 26.12, 422/140. If I am right that lawbreakers meet the Hobbesian definition of madness, does it follow that the sovereign can *never* punish criminals? Certainly not. The lawbreaker should be understood as a sane subject who momentarily succumbs to a fit of madness, which does not suffice to absolve him of legal responsibility. When he explicitly discusses "Mad-men", Hobbes evidently has in mind human beings whose reasoning is permanently disabled to a degree so significant that they do not even count as natural persons: see the discussion of personality in subchapter 1(a), especially the text corresponding to footnotes 18-20.
[207] As mentioned in subchapter 2(a), Hobbes carves out an exception when the subject must break some law to avoid "the terrour of present death". The subject ought then to be totally excused for his actions: *L*, 27.25, 468/157.
[208] *L*, 28.9, 484/162.

madman. But what of the criminal who predicts that his crime will go unpunished? Hobbes says he would have to be a "Foole" to act on his deviant desire.[209] Stated simply, Hobbes's argument is that disobedience subverts our long-term good. The only way to escape the horrors of the state of nature is to enter a commonwealth and obey the laws set down by our sovereign.[210] Even if a discrete instance of lawbreaking carries a low probability of producing anarchy, the potential cost is so great that it could never be rational to risk it. Natural reason ought to motivate us to give priority to our fear of violent death. We could only commit a crime if our reason failed to show us its true long-term consequences, or if it did so and we gave priority to some other passion. Both are marks of madness.

I pause to make three observations. First, the lawbreaker's problematic mental state bears a relation to imprudence.[211] Hobbes understands prudence as an intellectual virtue, grounded in experience, which allows human beings – and even certain animals – to predict the possible effects of their actions. In Hobbes's view, a person who commits an imprudent action falls into one of three categories: (1) he exercises prudence but commits an error; (2) he lacks the intellectual capacity to exercise prudence; or (3) he fails to exercise prudence despite possessing the requisite capacity. The first situation does not involve any defect in deliberation and it is debatable whether the resulting action should even be considered imprudent. However, the other situations involve faulty deliberation and therefore correspond to types of madness: the second to an inability to guide one's thoughts, the third to an inability to rationally control one's passions.

---

[209] See *L*, 15.4-7, 222-24/72-73. For an interesting debate about the nature of Hobbes's reply to this hypothetical fool, contrast Hoekstra (1997) to Hayes.
[210] *L*, 17.1, 254/85.
[211] *L*, 3.4-7, 40-43/9-10 and 5.5, 68/19 discusses prudence. See also the text corresponding to footnotes 170-73 and 191.

Second, Hobbes's lawbreaker bears a likeness to the schizophrenic of contemporary psychiatry, whose mind cannot properly connect feeling, thought, and action. When he deliberates, the lawbreaker is waylaid by incoherent thoughts generated by his feelings or, alternatively, by feelings so powerful that they supplant the thoughts he reflectively endorses.

Third, I previously claimed that the social contract includes distinct *authorisation* and *alienation* clauses; the latter produces an obligation to respect the sovereign's laws. Since it is unjust to breach an obligation, a subject commits an injustice at the moment he breaks the law.[212] Hobbes also says that "*Injury*, or *Injustice*, in the controversies of the world, is somewhat like to that, which in the disputations of Scholers is called *Absurdity*."[213] Absurdity is a form of madness. Does Hobbes make this comparison to suggest that the criminal is mad *because* he contradicts himself? I do not believe this conclusion follows. Although the criminal necessarily contradicts himself, some of these contradictions lie between healthy and mad wills, whereas others lie entirely between healthy wills. There is little textual evidence for the view that the latter contradictions amount to madness. Contradiction is an inescapable symptom of criminality: the criminal contradicts himself when he is mad *and* when he is sane.

A brief example can help show how this occurs. Imagine a man named John, whose fear of violent death causes him to alienate his right to self-government to the sovereign's benefit

[212] See the two relevant portions of subchapter 1(a): one pertains to Hobbes's conception of rights and obligations (footnotes 33-38 and the corresponding text), the other to the alienation clause in the social contract (footnotes 66-70 and the corresponding text).

[213] *L*, 14.7, 200-02/65. Kavka, 306-07, picks up on this passage but does not analyse it in any depth. In a recent investigation of Hobbes's theory of language, Duncan, 66, remarks that Hobbes draws this analogy because injustices and (some kinds of) absurdities involve contradiction. Although he does not examine this specific passage, Gauthier, 46, affirms that "Hobbes does not argue that an act is unjust *because* it is contradictory." Darwall (1995), 61-79, touches upon this theme when discussing how Hobbes changed some of his views on obligation in *Leviathan*. None of these authors considers the relationship between contradiction and mental illness.

(this flows from his first will, "W_1"). Later, the sovereign wills to prohibit robbery ("W_2") and promulgates a law to that effect. I recall that John owns W_2 because he authorised it. Even later, John deliberates about whether he should rob a bank. His greed prevails and he wills to commit robbery ("W_3"). W_3 is a mad will that John had no right to create after he obligated himself to obey the sovereign.[214] Moreover, it contradicts his prior healthy wills (W_1 and W_2).

And yet Hobbes's RSD introduces the possibility of contradiction between healthy wills. After John robs the bank, imagine that the sovereign discovers his crime and wills to punish him ("W_4"). John then finds out that the authorities are after him and wills to evade punishment ("W_5"). Both these wills arise from proper deliberation. A passion like fear of disorder can lead the sovereign to create W_4 without controversy. W_5 is also sound because natural reason counsels John to choose resistance over submission to punishment, at least according to what Hobbes says about the RSD. In addition, John is destined to continue producing healthy, contradictory wills until he is punished or the sovereign loses the capacity to punish him (for example, John might escape to a foreign country). I recall that a Hobbesian will is not a general mood or disposition. Rather, it is "*the last Appetite in Deliberating*,"[215] and deliberation continues "till the thing be either done, or thought impossible."[216] Accordingly, the content of a particular will is restricted to a particular "thing"; that will ceases to exist as soon as that "thing" is performed or can no longer be performed. Strictly speaking, then, the sovereign engages in

[214] This will is mad because it is always against natural reason to commit an injustice, as Hobbes makes clear in his reply to the fool: see footnote 209 and the corresponding text.
[215] *L*, 6.53, 92/28.
[216] *L*, 6.49, 90/28.

iterated deliberations motivated by a general desire to punish John.[217] Each deliberation culminates in a distinct will, which produces a punitive action which succeeds or fails. Similarly, John's iterated deliberations are motivated by a general desire to escape punishment. They culminate in various actions which tend towards that goal. Even though all of John's *authorised* and *personal* wills emerge from sound deliberation, they constantly contradict one another, up until the moment when he is punished or his punishment becomes impossible.

At this juncture, the criminal's ordeal resembles dissociative identity disorder, as he appears to bear two persons at once: a *private person* acting through his own body to resist punishment and a *public person* acting through representatives to inflict punishment on himself. When punishment enters the picture, the criminal is split between two personalities with coherent mental processes. Although contemporary psychiatry recognises this predicament as a disorder, it is *not* captured by Hobbes's account of madness, which is limited to a kind of schizophrenia.[218] Nonetheless, people with multiple personalities typically exhibit a single coherent personality within a bounded period of time; their various personalities succeed each other and are incoherent when compared amongst themselves. In contrast, Hobbes's criminal forms contradictory thoughts and performs contradictory actions *during the same period of time*. Unlike the schizophrenic, though, he does not suffer from a faulty transmission line between passion, reason, and action. Rather, he creates contradictory wills with regard to the same issue through independent, concurrent deliberations.

[217] The same passion can persist over a period of time, translating into different wills that produce different actions.
[218] For the definition of these disorders, see footnote 205 and the corresponding text.

The fact that Hobbes only considers the criminal to be mad during the first leg of his odyssey from crime to punishment engenders numerous problems to which I alluded at the outset of this chapter. To begin, it is simply implausible to claim that people who simultaneously will contradictory things are sane. Contemporary psychiatry has devised the category of "dissociative identity disorder" to describe this peculiar condition, which has long intrigued artists and philosophers.[219] On this point, Hobbes's psychological theory is not only deficient when compared to external standards; as I showed, it is also inconsistent with the claims he makes about absurdity in *Leviathan*.[220] Rather than extending his condemnation of absurdity to the fugitive's case, Hobbes turns sharply in the opposite direction, taking pains to insist that a subject may resist punishment reasonably and rightfully.[221] At first glance, it is quite baffling that Hobbes would choose to confer a normative stamp of approval to this dissociated state.

This brings us to a second shortcoming: Hobbes's refusal to call the fugitive insane destabilises his broader political theory. The legitimation of unlimited disobedience after a single act of illegitimate disobedience renders political obligation ephemeral. At least a few times in their lives, most people will meet the threshold of madness required to perform an illegal act which carries the possibility of imprisonment on paper (such as breaking the speed limit) – and yet Hobbes would only call them insane for driving too quickly, not for firing a gun at the cops who come to give them a ticket. Nor does this problem only affect subjects after they have

[219] Celebrated literary representations of this disorder include Robert Louis Stevenson's Dr. Jekyll and Mr. Hyde, as well as J. R. R. Tolkien's Gollum.
[220] See the quotation reproduced in the text corresponding to footnote 213; it is from *L*, 14.7, 200-202/65.
[221] As demonstrated by the discussion of natural reason throughout subchapter 2(b), Hobbes connects reason and right. Our reason counsels us to pursue our own good and we concomitantly have the right to do so. For a recent analysis of rationality, reasons of the good, and reasons of the right in Hobbes, see Abizadeh (2018), discussed in footnote 50.

broken the law: if people in the state of nature shared Hobbes's views on self-defence, they would predict this eventuality and would never agree to the social contract, thereby jeopardising the entire project of *Leviathan*. While these are serious problems, they need not be fatal. I will show that their germ is not to be found in anything Hobbes says about madness or rationality, but rather in his conception of self-defence, which affects his theory of madness derivatively. By reconstructing Hobbes's RSD, we can fortify his broader theory of politics.

d) Reconstructing Hobbes's Right to Self-Defence

This chapter has established (1) that the Hobbesian RSD allows for resistance to most forms of punishment and (2) that this in turn generates the paradox of the two-willed criminal. To be clear, the mere existence of a subject with two wills is not problematic; it is Hobbes's claim that a subject can hold two contradictory wills that are each formed through *sound deliberation* which disturbs his psychological and political theories. In this final subchapter, I tackle this problem from an indirect route. I propose to narrow the Hobbesian RSD so as to preclude resistance to most forms of punishment. If my move is successful, I will have eliminated the aforementioned paradox: in my reconstructed theory, one of the criminal's two conflicting wills – *i.e.* the will to resist punishment – could only arise through defective deliberation. The empirical possibility of a two-willed subject poses no special challenge if one of his wills is necessarily pathological; Hobbes's theory would then simply describe a form of madness as opposed to legitimating it.

My project rests on the assumption that the scope of the RSD in the text is broader than what is strictly necessary to meet Hobbes's normative commitments; or, stated differently, that Hobbes's rationale justifies a RSD which is narrower than the one he explicitly propounds. My

task, therefore, is to isolate this rationale and reformulate the RSD in a manner which meets its minimum requirements only. As I will show, the resulting right eliminates the problems outlined in this chapter all while fulfilling its intended function within Hobbes's political thought.

There can be no doubt that the RSD allows the subject to resist threats to his life, including the threat of capital punishment. The puzzle is why Hobbes additionally extends the right to corporal punishment and imprisonment.[222] An answer which immediately comes to mind is that Hobbes might have believed that a serious risk of death attached to these punishments, due to reasonable – yet historically contingent – assumptions about the techniques used to inflict pain and the conditions prevailing in prisons. The historical record provides ample evidence for this argument; for instance, Sreedhar (who ultimately rejects the argument) reproduces contemporary accounts of punitive practices in England during Hobbes's era. It is worth quoting from these accounts to elucidate the horrific backdrop to Hobbes's writing:

> The roge being apprehended, committed to prison, and tried in the next assises ... if he happen to be convicted for a vagabond, either by inquest of office, or the testimonie of two honest and credible witnesses upon their oths, he is then immediately adjudged to be greevouslie whipped and burned through the gristle of the right eare, with a hot iron of the compasse of an inch about, as a manifestation of his wicked life, and due punishment received for the same ... If he be taken a second time ... he shall be whipped againe, bored likewise through the other eare.[223]
>
> Beleve me, it greeveth me to heare (walking in the streetes) the pittifull cryes and miserable complayntes of poore prisoners in durance for debte, and the like so to continue for the rest of their life, destitute of libertie, meate, drink (though of the meanest sorte), and clothing to their backes, lying in filthie straw and lothsome dung, worse than anie dogge, voyde of all charitable consolation and brotherly comforte in the worlde, wishing and thirsting after deathe, to set them at libertie and loose them from their shackles ... and iron bandes.[224]

One could reasonably predict that mutilation, whipping inflicted "greevouslie", exposure to the elements without clothing, and confinement to quarters filled with "filthie strawe and

[222] See the passages listed in footnotes 105 and 107.
[223] Sreedhar, 61, citing Harrison, 272.
[224] Sreedhar, 67, citing Harrison, 287.

lothsome dung" might lead to death from disease or exhaustion, or at least a condition where one ends up "wishing and thirsting after deathe". The latter point is important: as Abizadeh writes,

> 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 the antonym of 'death'. [...] '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.[225]

Indeed, Hobbes is clear at various point in *Leviathan* that individuals seeking self-preservation aim for more than bare biological survival. He places "Desire of such things as are necessary to commodious living" alongside "Feare of Death" as a motive for the social contract,[226] whose members' "finall Cause, End, or Designe...is the foresight of their own preservation, and of a more contented life thereby";[227] grounds natural law "in the means of so preserving life, as not to be weary of it";[228] insists that subjects in civil society retain the right to "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";[229] and asserts that the sovereign's obligation to provide safety to the people does *not* mean "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".[230]

These passages invite two possible readings. On the first, *minimal* reading, Hobbes is placing a narrow category of situations on par with death: *i.e.* situations of extreme suffering with no foreseeable end which render life so miserable that it is not worth living for a rational individual. Consequently, the individual is entitled to use his RSD to avoid such situations just as he is entitled to avoid death itself. In practical terms, this extension of the RSD is rather modest

[225] Abizadeh (2018), 135 (see generally 131-38).
[226] *L*, 13.14, 196/63. (The quotes reproduced up to footnote 230, inclusively, are raised in Abizadeh (2018), 136.)
[227] *L*, 17.1, 254/85.
[228] *L*, 14.8, 202/66.
[229] *L*, 15.22, 234/77.
[230] *L*, 30.1, 520/175.

because the danger of falling into a condition as bad as death will likely arise very rarely in civil society. The second, *quality of life* reading has much wider ramifications. On this interpretation, individuals do not enter the social contract solely to insure against death and extreme suffering; they also wish to attain a sufficiently high standard of living. Therefore, individuals may disobey their sovereign if they are dissatisfied with their standard of living; the right to resist punitive sentences such as imprisonment is an appendage of this broader right.

Some scholars, such as Curran, unabashedly defend the quality of life reading.[231] Interestingly, while Sreedhar concedes that "quality of life considerations are not *on their own* sufficient to ground an act of justified resistance",[232] two of her three arguments for the inalienability of the RSD depend on quality of life premises – just like her justification for the right to rebel.[233] In what follows, I first demonstrate that Sreedhar's two best arguments collapse into the quality of life reading; next, that her third argument fails; finally, that the quality of life reading is flawed. I then argue that Hobbes's RSD coheres with the minimal reading and reconstruct the right accordingly.

Sreedhar claims that three principles justify resistance rights in Hobbes.[234] These principles render certain rights inalienable in the social contract specifically, and every subject can draw on this arsenal of reserved rights to resist authority. The first of these is the *reasonable expectations principle*, which holds that it is impossible for a contracting party to take on an obligation which he cannot reasonably be expected to perform. While the principle is valid, its

[231] Curran, 107-110.
[232] Sreedhar, 65-66.
[233] See footnotes 123-29 and the corresponding text.
[234] Sreedhar, 28-52. As stated previously, Sreedhar maintains that there are numerous independent resistance rights in Hobbes, while I maintain that there is a single right to resist death and extreme suffering.

results depend on the kinds of obligations Hobbes places outside the bounds of reasonable expectations. Quite obviously, Hobbes believes that a rational person cannot be expected to submit to death and allows him to resist it. However, Sreedhar further claims that this principle justifies resistance to corporal punishment or imprisonment which will not foreseeably cause death. Her comments on the matter oscillate between the minimal and quality of life readings. On the one hand, she asserts that most people cannot be expected to accept "great physical pain" or "a lifetime period of suffering and disfigurement";[235] this is consistent with the minimal reading and would *not* justify resistance to punishments which do not carry such risks. On the other hand, Sreedhar reads Hobbes as suggesting that "freedom from the kind of physical restraint imposed by chains and bars is essential to even a minimally acceptable, if not a good, life",[236] and so no person can be expected to submit to imprisonment in any form. This overblown claim evidently relies on quality of life concerns; it assumes that the parties to the social contract condition their consent on attaining much more than a life which is simply better than death.

Sreedhar's second principle, the *fidelity principle*, holds that rights cannot be transferred in a covenant if their transfer would contradict its very purpose. Even conceding the validity of the principle, its application depends on the purpose one ascribes to the social contract. On the minimal reading, individuals merely seek to escape the danger of death and unendurable suffering inherent to the state of nature; therefore, they cannot transfer their right to oppose such danger. Once again, however, Sreedhar hints at a more substantial purpose. From Hobbes's

[235] Sreedhar, 63-64.
[236] *Ibid.*, 68.

statements that people enter society in hope "of a more contented life", retaining the right to "all things else, without which a man cannot live, or not live well",[237] she infers that their objective is to avoid *all* violence and loss of liberty, and then concludes that they retain the right to resist any form of corporal punishment or imprisonment.[238] As such, the success of the argument depends on the success of the quality of life reading.

Before proceeding to the merits of that reading, it is worth examining Sreedhar's final principle, which does not depend upon it. The *necessity principle* holds that the parties to the social contract cannot alienate rights which are unnecessary to establish an effective government. Sreedhar acknowledges that "[the] necessity principle is not, strictly speaking, a principle of valid covenants. Hobbes does not say that unnecessary transfers of rights are thereby invalid."[239] However, Hobbes does insist that subjects' retention of the RSD does not significantly impair the sovereign's punitive capacity; Sreedhar seizes upon these claims to argue that this principle operates exceptionally in the social contract. The argument is unwarranted. Hobbes's claims are intended to assuage concerns that the RSD, which he justifies on *different grounds*, might undermine the power of the state; they do *not* provide an independent ground for the right, nor do they delineate its scope. Sreedhar fails to establish either that this special principle of covenants applies to the social contract or that considerations of necessity ever hold analytical relevance for the alienability of rights.

[237] *L*, 15.22, 234/77 and 17.1, 254/85.
[238] Sreedhar, 64 and 68-69, referring to the aforementioned passages and several others.
[239] *Ibid.*, 51.

I now return to the quality of life reading and evaluate its strength. At least some of the passages upon which its proponents rely do establish that individuals enter civil society with the intention of enjoying a life which is not only endurable but also as happy as possible. When describing the state of nature, Hobbes writes:

> In such condition, there is no place for Industry; because the fruit thereof is uncertain: and consequently no Culture of the Earth; no Navigation, nor use of the commodities that may be imported by Sea; no commodious Building; no Instruments of moving, and removing such things as require much force; no Knowledge of the face of the Earth; no account of Time; no Arts; no Letters...[240]

This knowledge can be imputed to the inhabitants of the state of nature, who can reasonably hope to reap the fruits of industry, agriculture, architecture, literature, and the like once they have left their miserable condition.

But do subjects retain the right to second-guess their sovereign and forcefully resist him if they are dissatisfied with their standard of living? It is this stronger, much more contestable claim that lies at the heart of the quality of life reading. This reading is flawed because it elevates the subject's private judgement over the sovereign's public judgement. Throughout *Leviathan*, Hobbes hammers home the message that individuals acting on their private judgement risk disagreeing, fighting, and killing each other.[241] The Hobbesian social contract solves this problem by subordinating the judgement of every individual to that of a public official, the sovereign, who sets common rules and standards. Hobbes makes this claim clearly and it is not seriously disputed in the literature.[242] And so even if individuals enter the social contract to obtain "all other Contentments of life, which every man by lawfull Industry, without danger, or

---

[240] *L*, 13.9, 192/63.
[241] See *e.g. L*, 5.3, 18-19, cited in footnote 31.
[242] When explaining the social contract, Hobbes asserts that the parties agree to "submit their Wills, every one to [the sovereign's] Will, and their Judgements, to his Judgment": *L*, 17.13, 260/87.

hurt to the Common-wealth, shall acquire to himselfe",[243] it is the sovereign who ultimately must decide what kinds of "Contentments" his subjects may pursue – and how they may do so – without imperiling the body politic. As such, Sreedhar's *fidelity principle* is problematic to the extent it assumes that subjects retain the right to evaluate whether the social contract has achieved its goal of sufficiently raising their standard of living and to act on those evaluations. Persons in the state of nature agree that the best means for improving their quality of life is to respect the judgement of the sovereign. Therefore, acts of disobedience or rebellion motivated by political disagreement undermine the intended operation, and hence the overarching purpose, of the social contract.

By granting subjects the RSD, though, Hobbes intentionally declines to wholly eliminate private judgement in civil society. What must be kept in mind is that the scope of the RSD is directly proportional to the scope of permissible private judgement. All things being equal, a narrow interpretation of the RSD is preferable to a broader one because it promotes greater political stability. Hobbes is unconcerned with the residue of private judgement implicated by the RSD because he adheres to the *minimal* view: this right is opposable to direct threats of death or extreme suffering only. Such a narrow right is unlikely to cause significant problems in a commonwealth for two reasons. At the empirical level, we can plausibly assume that individuals in a functional society will rarely perceive threats of death or unendurable suffering, while frequently doubting the efficacy of government policies. At the epistemic level, it is easier for rational individuals to make a correct judgement about whether they face an immediate danger of death or unendurable suffering than whether the sovereign is providing an adequate standard of

[243] *L*, 30.1, 520/175.

living. Consequently, individuals acting upon their private judgement within this narrow zone will rarely decide to oppose their sovereign. The same logic applies to punishment: the Hobbesian subject is only justified in resisting his sentence if he privately reasons that it will likely cause death or extreme suffering. Otherwise, if the sovereign decides that he must punish a subject *and* the subject reasons that his punishment will not lead to such consequences, the subject remains under his general political obligation of obedience.

At this juncture, I should clarify that I include serious, ongoing psychological agony under the rubric of extreme suffering. Hobbes declares that the subject may legitimately refuse to incriminate a person "by whose Condemnation a man falls into misery; [such as] a Father, Wife, or Benefactor"[244], or to perform a "dishonourable Office".[245] While Hobbes does not provide an example for the latter claim in *Leviathan*, in *De Cive* he specifies that a subject may resist a command to execute his own father, since "a son may prefer to die rather than live in infamy and loathing".[246] Hobbes might be relying on a timeless truth about the agony of a guilty conscience or on cultural assumptions about honour specific to his era; in either case, he does not significantly expand the scope of the RSD and permissible private judgement. Orders to dishonour oneself in a manner which entails enduring psychological misery ought to be both rare and easy to identify.[247]

---

[244] *L*, 14.30, 214/70.
[245] *L*, 21.15, 338/113. Hobbes also includes "dangerous" offices within this branch of the RSD, but this is easily explained by reference to the danger of death or extreme physical suffering.
[246] *De Cive*, chapter 6, section 13, cited in Sreedhar, 77.
[247] Despite permitting subjects to rebut the presumption of obedience by privately judging that an order is dishonourable, Hobbes asserts that they can disobey only if they make a further private judgement that disobedience will not endanger society: *L*, 21.15, 338/112, discussed in footnotes 111-13 and the corresponding text. Hobbes

I have demonstrated that Hobbes's arguments only justify what I have called a minimal RSD; his explicit extension of the right beyond this boundary relies on empirical assumptions which do not hold universally. To conclude this subchapter, I will demonstrate that reconstructing the RSD in accordance with its philosophical justification has beneficial implications for Hobbes's theory of madness. My proposal is simple: within the context of the criminal law, the RSD ought to be reformulated so as to only permit subjects to resist punishment carrying a reasonable risk of death or extreme suffering. This proposal releases the Hobbesian RSD from its historical moorings while respecting its underlying rationale.

The upshot is that the criminal loses the right to resist most forms of punishment, including the most common punishment imposed for serious crimes in contemporary liberal democracies: imprisonment. A rational agent concerned with his own self-preservation as understood by Hobbes – *i.e.* as the avoidance of death or a life worse than death – ought to calculate that submission to a provisional term of imprisonment is preferable to resistance, for after experiencing the temporary harm of punishment, he will recover the full enjoyments of society. Even submission to a life sentence is rationally preferable to violent resistance (with the attendant danger of serious injury or death), provided that carceral conditions are not so awful as to appear as bad as death itself.[248] The will to resist these punishments could *only* emanate from unsound deliberation, since the criminal would thereby fail to privilege his desire for self-

decreases the likelihood that subjects will disobey by assigning two competing questions to their private judegment. As such, Sreedhar's reliance on this passage to carve out a large space for disobedience is unwarranted: 78-81.

[248] It is an open question whether the actual conditions prevailing in many prisons are more attractive than a life on the run, especially considering the prevalence of violence between inmates, abuse at the hands of guards, and the employment of disciplinary methods such as solitary confinement. Nevertheless, the case I am making only fails if conditions are clearly so bad that a rational agent would prefer to face the risk of death. The fact that most criminals do not violently resist arrest or attempt to break out of jail stands as presumptive, though rebuttable, evidence that they are not so bad.

preservation. The lawbreaker-turned-fugitive would be continuously "mad" throughout his journey: not because he holds contradictory wills, but rather because some of the wills responsible for his particular contradictions would necessarily have a defective origin. The unfortunate consequences of the RSD for the Hobbesian criminal's psychology are thus nullified in commonwealths which forbid the death penalty and the infliction of unendurable, long-lasting suffering.[249]

Conclusion

In this thesis, I hope to have disentangled conceptual knots introduced by the RSD into Hobbes's account of punishment along two dimensions: the *origin* and *execution* of punishment. I began the chapter on the origin of punishment by explaining three foundational concepts in Hobbes's legal theory: personality, authorisation, and the alienation of rights. Next, I defended an interpretation of the social contract as a two-step procedure: (1) first, every subject enters a covenant with every other subject, thereby *authorising* the sovereign to act in the name of each one; (2) second, every subject *alienates* a portion of his natural right as a gift to the sovereign while retaining a residue – the RSD. When the sovereign punishes a subject, he exercises *his own* natural right to violence but does so pursuant to each subject's authorisation. On this model, every subject has authorised the punishment of every subject – even himself, and even if he should choose to exercise his residual RSD whenever the sovereign attempts to punish him. Hobbesian punishment is thus an institution of public law.

[249] For discussion of the death penalty in Hobbes, see Heyd. Numerous liberal democracies have legal prohibitions on "cruel and unusual" punishment; whether they always comply with those prohibitions is a different matter.

Nevertheless, the retained RSD creates problems for the execution of punishment. I opened the second chapter with an investigation of the scope of the RSD, concluding that it permits resistance to most forms of punishment (including capital punishment, corporal punishment, and imprisonment). I argued that such a right has deleterious pragmatic and conceptual consequences for Hobbes's political system. Pragmatically, it weakens the permanence of political obligation, increases the likelihood of violence, and renders the social contract unappealing to rational agents. Conceptually, it produces the paradox of the criminal with *two contradictory wills*: his authorised will to punish himself and his personal will to avoid punishment. To examine this incongruous figure, I developed an *original account of Hobbesian madness* which captures four classes of people who deliberate unsoundly: (1) those with passions of abnormal intensity; (2) those with passions with abnormal content; (3) those with unguided passions; and (4) those with unguided or absurd thoughts. I demonstrated that the criminal is "mad" when he breaks the law but not when he resists punishment, and that this is a problematic diagnosis. I concluded with a solution: if the RSD were reconstructed so as to fit only the minimum requirements of Hobbes's justification – *i.e.* the pursuit of self-preservation understood as *the avoidance of death or extreme suffering* – then resistance to many common forms of punishment, such as imprisonment, becomes illegitimate. A criminal who resists the authorities would then necessarily hold a pathological will and fall under the rubric of Hobbesian madness.

This thesis has not come close to exhausting the full range of questions raised about punishment in Hobbes's rich political theory. For example, the conception of madness I derived from and applied to Hobbes's theory of punishment would benefit from engagement with his other works and competing contemporary conceptions, such as the medieval theory of the

humours. On another note, the RSD invites doubts about the *purpose* of punishment; in particular, it appears to indicate scepticism about its corrective potential, even though Hobbes works this function into his formal definition.[250] The Hobbesian idea of authorisation also foreshadows modern retributivist theories by insisting that the criminal wills to punish himself. Unlike the mature retributivists, though, it is unclear whether Hobbes believes that punishment gives criminals their moral deserts; indeed, the RSD might suggest otherwise. But these questions – and many more – must be left for another day.

[250] *L*, 28.1, 481/161 and 28.7, 484/162. In his formal definition, Hobbes states that punishment must be inflicted "to the end that the will of men may thereby the better be disposed to obedience"; he later clarifies that "evil 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." He thus provides two consequentialist justifications for punishment: correction and deterrence.

Bibliography

Hobbes's Works

Hobbes, Thomas. 1997 [1642, 1647]. *De Cive*. Edited by Richard Tuck and Michael Silverthorne. Cambridge, UK; Cambridge University Press.

—. 2012 [1651, 1668]. *Leviathan: The English and Latin Texts*. Edited by Noel Malcolm. Oxford: Clarendon Press.

Other Primary Texts

Bramhall, John. 1995 [1658]. "The Catching of Leviathan, Or the Great Whale." In *Leviathan: Contemporary Responses to the Political Theory of Thomas Hobbes*, edited by G. A. J. Rogers, 115-80. Bristol: Thoemmes Press.

Justinian. 1970 [533]. *Institutes*. Edited by Thomas Collett Sandars. Westport, CT: Greenwood.

Locke, John. 1988 [1689]. *The Second Treatise of Government*. In *Two Treatises of Government*, edited by Peter Laslett, 265-428. Cambridge, UK: Cambridge University Press.

Rousseau, Jean-Jacques. 1997 [1762]. *Of the Social Contract*. In *The Social Contract and other later political writings*, edited by Victor Gourevitch, 39-152. Cambridge, UK: Cambridge University Press.

Secondary Texts

Abizadeh, Arash. 2016. "Sovereign Jurisdiction, Territorial Rights, and Membership in Hobbes." In *The Oxford Handbook of Hobbes*, edited by A. P. Martinich and Kinch Hoekstra, 397-431. Oxford: Oxford University Press.

—. 2017. "Hobbes on Mind: Practical Deliberation, Reasoning, and Language." *Journal of the History of Philosophy* 55 (1): 1-34.

—. 2018. *Hobbes and the Two Faces of Ethics*. Cambridge, UK: Cambridge University Press.

Barry, Brian. 1972. "Warrender and His Critics." In *Hobbes and Rousseau: A Collection of Critical Essays*, edited by Maurice Cranston and Richard S. Peters, 37-65. New York: Doubleday.

Baumgold, Deborah. 1988. *Hobbes's Political Theory*. Cambridge, UK: Cambridge University Press.

Blau, Adrian. 2016. "Reason, Deliberation, and the Passions." In *The Oxford Handbook of Hobbes*, edited by A. P. Martinich and Kinch Hoekstra, 195-220. Oxford: Oxford University Press.

Burgess, Glenn. 1994. "On Hobbesian Resistance Theory." *Political Studies* 42 (1): 62-83.

Cohen, Andrew. 1998. “Retained Liberties and Absolute Hobbesian Authorization.” *Hobbes Studies* 11 (1): 33-45.

Copp, David. 1980. “Hobbes on Artificial Persons and Collective Actions.” *Philosophical Review* 89 (4): 579-606.

Cromartie, Alan. 2011. “*The Elements* and Hobbesian Moral Thinking.” *History of Political Thought* 32 (1): 21-47.

Curran, Eleanor. 2007. *Reclaiming the Rights of the Hobbesian Subject*. New York: Palgrave Macmillan.

Darwall, Stephen L. 1995. *The British Moralists and the Internal “Ought”: 1640-1670*. Cambridge, UK: Cambridge University Press.

—. 2000. “Normativity and Projection in Hobbes’s *Leviathan*.” *The Philosophical Review* 109 (3): 313-47.

Dedek, Helge and Martin J. Schermaier. 2012. “Obligation, Greek and Roman.” In *The Encyclopedia of Ancient History*, available online: <https://doi-org.proxy3.library.mcgill.ca/10.1002/9781444338386.wbeah13191> (January 1, 2019).

Deigh, John. 1996. “Reason and Ethics in Hobbes's *Leviathan*.” *Journal of the History of* Philosophy 34 (1): 33-60.

Duncan, Stewart. 2016. “Hobbes on Language: Propositions, Truth, and Absurdity.” In *The Oxford Handbook of Hobbes*, edited by A. P. Martinich and Kinch Hoekstra, 60-75. Oxford: Oxford University Press.

Dyzenhaus, David. 2001. “Hobbes and the Legitimacy of Law.” *Law and Philosophy* 20 (5): 461-91.

—. 2010. “Hobbes’s Constitutional Theory.” In *Leviathan*, edited by Ian Shapiro, 453-80. New Haven: Yale University Press.

—. 2012. “Hobbes on the Authority of Law.” In *Hobbes and the Law*, edited by David Dyzenhaus and Thomas Poole, 186-209. Cambridge, UK: Cambridge University Press.

Fox-Decent, Evan. 2012. “Hobbes’s Relational Theory: Beneath Power and Consent.” In *Hobbes and the Law*, edited by David Dyzenhaus and Thomas Poole, 118-44. Cambridge, UK: Cambridge University Press.

Frost, Samantha. 2008. *Lessons from a Materialist Thinker: Hobbesian Reflections on Ethics and Politics*. Stanford: Stanford University Press.

Gauthier, David. 1969. *The Logic of Leviathan: The Moral and Political Theory of Thomas Hobbes*. Oxford: Oxford University Press.

Gert, Bernard. 2001. “Hobbes on Reason.” *Pacific Philosophical Quarterly* 82 (3-4): 243-57.

Gordley, James. 1991. *The Philosophical Origins of Modern Contract Doctrine*. New York: Oxford University Press.

Green, Michael J. 2015. "Authorization and Political Authority in Hobbes." *Journal of the History of Philosophy* 53 (1): 25-47.

Hampton, Jean. 1986. *Hobbes and the Social Contract Tradition*. Cambridge, UK: Cambridge University Press.

Harrison, Molly. 1962. *How They Lived*. Oxford: Blackwell.

Hayes, Peter. 1999. "Hobbes's Silent Fool: A Response to Hoekstra." *Political Theory* 27 (2): 225-29.

Heyd, David. 1991. "Hobbes on Capital Punishment." *History of Philosophy Quarterly* 8 (2): 119-34.

Hoekstra, Kinch. 1997. "Hobbes and the Foole." *Political Theory* 25 (5): 620-54.

Hoekstra, Kinch. 2003. "Hobbes on Law, Nature, and Reason." *Journal of the History of Philosophy* 41 (1): 111-20.

Irwin, Terence. 2008. *The Development of Ethics, Volume 2: From Suarez to Rousseau*. Oxford: Oxford University Press.

Kavka, Gregory S. 1986. *Hobbesian Moral and Political Theory*. Princeton: Princeton University Press.

Lee, Daniel. 2012. "Hobbes and the Civil Law: The Use of Roman Law in Hobbes's Civil Science." In *Hobbes and the Law*, edited by David Dyzenhaus and Thomas Poole, 210-35. Cambridge, UK: Cambridge University Press.

Martinich, A. P. 2016. "Authorization and Representation in Hobbes's *Leviathan*." In *The Oxford Handbook of Hobbes*, edited by A. P. Martinich and Kinch Hoekstra, 315-38. Oxford: Oxford University Press.

May, Larry. 1992. "Hobbes on Fidelity to Law." *Hobbes Studies* 5 (1): 77-89.

Nagel, Thomas. 1959. "Hobbes's Concept of Obligation." *The Philosophical Review* 68 (1): 68-83.

Noll, Richard. 2007. *The Encyclopedia of Schizophrenia and Other Psychotic Disorders*. New York: Facts on File.

Norrie, Alan. 1984. "Thomas Hobbes and the Philosophy of Punishment." *Law and Philosophy* 3 (2): 299-320.

Pink, Thomas. 2003. “Suarez, Hobbes and the scholastic tradition in action theory.” In *The Will and Human Action: From Antiquity to the Present Day*, edited by Thomas Pink and M. W. F. Stone, 127-53. London: Routledge.

—. 2011a. “Thomas Hobbes.” In *A Companion to the Philosophy of Action*, edited by C. Sandis and T. O’Connor, 473-80. Oxford: Blackwell.

—. 2011b. “Thomas Hobbes and the Ethics of Freedom.” *Inquiry* 54 (5): 541-63.

—. 2016. “Hobbes on Liberty, Action, and Free Will.” In *The Oxford Handbook of Hobbes*, edited by A. P. Martinich and Kinch Hoekstra, 171-94. Oxford: Oxford University Press.

Poole, Thomas. 2012. “Hobbes on Law and Prerogative.” In *Hobbes and the Law*, edited by David Dyzenhaus and Thomas Poole, 68-96. Cambridge, UK: Cambridge University Press.

Riley, Patrick. 1982. *Will and Political Legitimacy: A Critical Exposition of Social Contract Theory in Hobbes, Locke, Rousseau, Kant, and Hegel*. Cambridge, MA: Harvard University Press.

Ristroph, Alice. 2012. “Criminal Law for Humans.” In *Hobbes and the Law*, edited by David Dyzenhaus and Thomas Poole, 97-117. Cambridge, UK: Cambridge University Press.

Runciman, David. 2000. “What Kind of Person Is Hobbes’s State? A Reply to Skinner.” *The Journal of Political Philosophy* 8 (2): 268-78.

Schrock, Thomas S. 1991. “The Rights to Punish and Resist Punishment in Hobbes’s *Leviathan*.” *The Western Political Quarterly* 44 (4): 853-90.

Skinner, Quentin. 2002. *Visions of Politics, Volume 3: Hobbes and Civil Science*. Cambridge, UK: Cambridge University Press.

Sreedhar, Susanne. 2010. *Hobbes on Resistance: Defying the Leviathan*. Cambridge, UK: Cambridge University Press.

Statt, David. 1998. *The Concise Dictionary of Psychology*. New York: Routledge.

Steinberger, Peter J. 2002. “Hobbesian Resistance.” *American Journal of Political Science* 46 (4): 856-65.

Taylor, Alfred Edward. 1938. “The Ethical Doctrine of Hobbes.” *Philosophy* 13 (4): 406-24.

van Apeldoorn, Laurens. 2012. “Reconsidering Hobbes’s Account of Practical Deliberation.” *Hobbes Studies* 25 (2): 143-65.

—. 2014. “Rationality and Freedom in Hobbes’s Theory of Action.” *History of European Ideas* 40 (5): 603-21.

Warrender, Howard. 1957. *The Political Philosophy of Hobbes: His Theory of Obligation*. Oxford: Oxford University Press.

Weber, Dominique. 2010. ""Cela n'équivaut pas à dire que les enfants et les fous sont privés de la liberté véritable": Hobbes et le problème de la folie." *Dix-septième siècle* 247 (2): 223-34.

Yates, Arthur. 2014. "The Right to Punish in Thomas Hobbes's Leviathan." *Journal of the History of Philosophy* 52 (2): 233-54.

Zimmerman, Reinhard. 1996. *The Law of Obligations: Roman Foundations of the Civilian Tradition*. Oxford: Clarendon Press.