

Volume XII, No. 1

Spring 2013

LYCEUM

The Surreptitious *Leviathan*:
Concealing the Beast of Scientific Reason
David W. Cheely

The Consensus View's Two Principles of Political Authority:
Pauline and Transmission
Leonard Ferry

Extended Mind and the Extension of a Self
Sangkyu Shin

Two Sorts of Self-Creation:
On Galen Strawson's "Basic Argument"
Griffin Klemick

**A Publication of the
Saint Anselm College Philosophy Department**

LYCEUM

The [LYCEUM](#) is a publication of the
[Saint Anselm College Philosophy Department](#).

Editors

Kyle Hubbard
Saint Anselm College

Joshua Tepley
Saint Anselm College

Board of Editorial Consultants

Robert Anderson
Saint Anselm College

Robert Augros
Saint Anselm College

David Banach
Saint Anselm College

Martha Catherine Beck
Lyon College

Montague Brown
Saint Anselm College

Drew Dalton
Florida Southern College

Fr. John Fortin, OSB
Saint Anselm College

Susan Krantz Gabriel
Saint Anselm College

Sarah Glenn
Saint Anselm College

Peter Josephson
Saint Anselm College

Thomas Larson
Saint Anselm College

Max Latona
Saint Anselm College

James Mahoney
Saint Anselm College

James O'Rourke
Saint Anselm College

Joseph Spoerl
Saint Anselm College

Kevin Staley
Saint Anselm College

Christopher Tollefsen
University of South Carolina

LYCEUM

Table of Contents

The Surreptitious <i>Leviathan</i> : Concealing the Beast of Scientific Reason <i>David W. Cheely</i>	1
The Consensus View's Two Principles of Political Authority: Pauline and Transmission <i>Leonard Ferry</i>	27
Extended Mind and the Extension of a Self <i>Sangkyu Shin</i>	82
Two Sorts of Self-Creation: On Galen Strawson's "Basic Argument" <i>Griffin Klemick</i>	101
About the Authors	126

The Surreptitious *Leviathan*: Concealing the Beast of Scientific Reason

David W. Cheely

Typically, in the political scholarship on Thomas Hobbes, the political theory expressed in the *Leviathan* is understood as logically independent of Hobbes' mechanistic materialism and his geometric methodology. Scholars intent upon viewing Hobbes' *magnum opus* strictly as a work of political theory identify the 'right of nature' as a first principle of his political philosophy, yet in doing so, they pass over, or de-emphasize the unity between this principle as it appears in the *Leviathan* and Hobbes' mechanistic view of human nature explicit in his earlier works. Furthermore, those scholars completely miss an important stylistic aspect of the *Leviathan*, namely its surreptitious tone. In this paper, I argue in support of the unity thesis and highlight two important consequences of viewing Hobbes' political theory as a consequence of his mechanistic view of human nature. Not only does the "grand atheist" accusation gain more traction, but the degree to which Hobbes' thought surreptitiously fools the majority of his contemporary readers can be better understood through a surreptitious reading of the *Leviathan*.

The Unity Thesis: Hobbes' Political Philosophy as a Consequence of His First Philosophy

We shall approach Hobbes' *Leviathan* through the unity thesis. It is my contention that Hobbes' political philosophy is an extension of, or an application of, his mechanistic first philosophy and his theory of knowledge. The best way to understand Hobbes' political philosophy, and to a lesser extent his theology, is to view Hobbes' philosophy as a unified whole. Hobbes begins the *Leviathan* with a description of the genesis and function of sense perception, a fact that Cees Leijenhorst (2007) finds quite surprising for a work that introduces Hobbes'

The Surreptitious Leviathan: Concealing the Beast of Scientific Reason

political theory and its implications for theology.¹ Hobbes' mechanical explanation of sense perception in the first two chapters of *Leviathan* are admittedly crude; however, as Leijenhorst (2007) writes in reference to *De Corpore*, "we do indeed find a more refined and extensive account [of sense perception] than offered in *Leviathan*" (p. 87). Though Leijenhorst's interest is in establishing the extent to which Hobbes' polemics in the first two chapters of *Leviathan* are the result of his anti-Aristotelian philosophy of science, his insight here is helpful toward determining whether or not Hobbes' methodology in *Leviathan* is consistent with his methodology in *De Corpore* and *The Elements of Law*.

Hobbes' methodology is akin to the geometric method in which complex theorems are derived from axiomatic first principles.² While In Switzerland in 1628, the forty year-old Hobbes discovered Euclid in a private library in Geneva.³ His discovery of Euclid prompted Hobbes to pursue philosophic truth through the method of geometry, where truth is a function of a reasoning that avoids contradiction through systematic analytical ordering. Later, when Hobbes would develop a mechanistic conception of the universe consisting only of extended body, it was the method of geometry, as observed by Hobbes in his discovery of Euclid, that provided the type of verbally acute lucidity necessary to

¹ References to *De Corpore* and *The Elements of Law (Human Nature and De Corpore Politico)* are derived from the *Oxford World's Classics (OWC)* edition listed in the end reference to this paper. This edition of Hobbes' works includes complete translations of *The Elements of Law*, along with three appended chapters from *De Corpore* (Chapters I, VI, and XXV). When citing from *De Corpore*, I will use the abbreviation "DCo" followed by the pagination from *OWC*. When citing from *The Elements of Law*, I will use the abbreviation "EL" followed by the pagination from *OWC*. References to *Leviathan*, will use the abbreviation "L" followed by the pagination from the Curley edition (1994) of *Leviathan*, also referenced in full in the end references to this paper.

² Jesseph (2002) expresses the same point when he writes, "Hobbes conceived of philosophy as an enterprise that begins with quite general definitions and principles and proceeds *more geometric* through demonstrative syllogisms to establish irrefutable truths" (p. 163).

³ Mintz (1962) presents a charming description of Hobbes' discovery and subsequent fascination with the proof method of geometry.

LYCEUM

present his mechanistic materialism as a systematic whole. In 1636, eight years after his exposure to Euclid, Hobbes, in a letter addressed to William Cavendish, writes of his displeasure with Walter Warner and Claude Mydorge for their failure to properly demonstrate their findings in optics. Hobbes (1994a) writes, “but they do not well to call their writings demonstrations, for the grounds and suppositions they use, so many of them as concerne light, are uncertayne and many of them not true” (*Correspondence*, p. 34). It is evident that Hobbes has a clear conception of the necessity of employing true first principles as a proper ground for one’s inquiry, if one’s inquiry is to yield accurate experimental results.

Hobbes’ first formulation of his mechanistic materialism can be found in an untitled work that was added as an appendix to Ferdinand Tönnies edition of Hobbes’ *The Elements of Law*.⁴ This work, roughly dated in 1637, presents Hobbes’ mechanistic materialism through a thoroughly deductive argument, in which conclusions are syllogistically drawn out from a few first principles or definitions. In section I (of III sections), Hobbes identifies four primary principles or definitions:

1. That, whereto nothing is added, and from which nothing is taken, remains in the same state it was.
 2. That which is in no way touch’d by another, hath nothing added to nor taken from it.
 3. Agent is that which hath power to move.
 4. [The patient is] that which hath power to be moved.
- (Brandt, 1927, p. 13)

Hobbes felt these four principles were self-evident, and were therefore certain. From these principles, Hobbes deduces the various laws of his mechanistic

⁴ See Brandt (1927) for a more detailed discussion of Hobbes’ *The Little Treatise*.

The Surreptitious Leviathan: Concealing the Beast of Scientific Reason

philosophy.⁵ These laws are meant to explain phenomena as experienced in nature; Hobbes feels sense experience can be explained through local motion alone. In short, if there is no motion, then nothing can be sensed. Brandt (1927) writes, “If we are to characterize the little treatise in one single sentence it will be in the following words: *The little treatise concerns the problem of the act of sense in a broad signification. Hobbes’ leading thought is that both the external and the internal act of sense and the higher, psychic functions derived therefrom are local motion and nothing but local motion*” (p. 46) Here, we see that Hobbes has remained true to his geometric method, deriving explanations for various sense experiences from first principles; each principle reflects both a generalization from empirical observations and a mechanism that requires bodies in motion for sense experience to occur. Recalling his earlier criticisms of Warner and Mydorge, we can see that Hobbes, the naturalist philosopher, has developed a systematic method for explaining phenomena that is consistent with his vision of the correct manner in which one ought to do philosophy. Furthermore, as we shall soon see, Hobbes’ methodology and mechanistic materialism has been refined so as to closely resemble the mechanism and methodological analysis presented in *De Corpore* and *The Elements of Law*.

Thus far, Hobbes’ strict mechanism seems to only apply to what we might call, today, physics, or neuroscience if one wishes to stretch things a bit. *De Corpore* and *The Elements of Law* succeed in bringing Hobbes’ mechanistic materialism and his geometric method to bear on his epistemology and theory of mind. Hobbes distinguishes between two types of knowledge—knowledge acquired through prudence and knowledge acquired through ratiocination, while claiming that prudence “is not to be esteemed philosophy” (*DCo*, 1999, p. 186). Hobbes here makes a case for a computational theory of mind, while still holding true to his empirical roots. In distinguishing between two types of knowledge, and insisting that philosophy should focus upon ratiocination, or computation,

⁵ Again, for reasons of brevity, we shall not go into the detail required to state and elucidate these laws as Hobbes’ describes them in 1637—I refer the reader to Brandt (1927).

LYCEUM

Hobbes at the same time distinguishes man from the brutes while also rejecting any innate hypothesis or any immaterial substantiality. Like the brutes, man extrapolates from past experiences, but unlike the brutes, man can demonstrate, through powers of ratiocination, truths that are the proper object of philosophic study. Hobbes writes, “we must remember and acknowledge that there be in our minds continually certain images or conceptions of the things without us, insomuch that if a man could be alive, and all the rest of the world annihilated, he should nevertheless retain the image thereof” (*EL*, 1999, p. 22). All mental images and concepts are derived from experience for Hobbes. Again, Hobbes holds to his empirical roots, while maintaining his distinction between the two types of knowledge when he writes, “and all experience being (as I have said) but remembrance, all knowledge is remembrance...the registers of the latter are called the sciences” (*EL*, p. 40). There are no other sources of knowledge besides perception, yet the second type of knowledge is distinct from the first type of knowledge despite beginning with the same raw materials. Human reason as Hobbes’ describes it, is the creation of man, made possible through language, which, through naming, identifies the concepts that makes this second kind of knowledge possible.⁶ Hobbes writes, “By the advantage of names it is that we are capable of science, which beasts, for want of them, are not” (*EL*, p. 36). Hobbes is a nominalist; he rejects the “universality of one name to many things” as that which “hath been the cause that men think the things themselves are universal” (*EL*, p. 36). Proofs manipulate words, not abstract objects; this manipulation is a systematic connection of words to propositions to syllogistic conclusions. Yet, this is not enough for one to profess scientific knowledge—one must also have evidence, which Hobbes defines as “the concomitance of a man’s

⁶ I use the word ‘creation’ in this sentence to coincide with my claim that the Leviathan is Reason as Hobbes’ conceives ratiocination, but the word ‘discover’ may, indeed, be more apt in terms of Hobbes’ philosophy, particularly considering his own discovery of Euclid. I am unwilling to submit Hobbes’ philosophy to the sort of anachronistically revisionist-historical analysis which would suggest a Darwinian explanation for the ‘evolutional appearing’ of Hobbes’ second type of knowledge, thus I have decided to stick with my rather misleading description of the genesis of the Leviathan as a ‘creation of man.’

The Surreptitious Leviathan: Concealing the Beast of Scientific Reason

conception with words that signify such conception in the act of ratiocination” (*EL*, p. 41). Though this may read like the writing of a rationalist philosopher, Hobbes’ insistence that both kinds of knowledge derive from experience suggests that Hobbes’ conception of scientific knowledge is grounded in experience, and that the distinctly human ability to manipulate words, propositions, and conclusions into the form of a truth claim, while also maintaining the ability to judge the appropriateness of evidence with respect to these truth claims, is itself explainable in terms of his mechanistic philosophy. It is this latter point which brings us to Hobbes’ theory of mind.

Hobbes believes in the transparency of the mind; there is no unconscious iceberg of mental casuistry beneath the surface of man’s reflective awareness. Furthermore, Hobbes’ strict mechanism and its deduced laws, i.e., that sensation requires bodies in motion, is extended to the mind—there is nothing immaterial about the mind which cannot be reduced to a science based upon the laws of cause and effect as they pertain to motion. Hobbes writes, “That the said image or colour is but an apparition unto us of that motion, agitation, or alteration, which the object worketh in the brain or spirits, or some internal substance in the head” (*EL*, p. 23). It is clear from this principle that Hobbes has retained the mechanism of bodies in motion that he describes in his *Little Treatise*. However, here Hobbes explicitly applies his mechanism toward understanding the manner in which the human mind functions. Brandt (1927) echoes this point when he writes, “The reason why moral philosophy is treated after physics is that the motions of the mind (as examples of such Hobbes mentions appetite, aversion, love, benevolence, hope, fear, emulation, and envy) have their causes in sense and imagination which are the subject of physical contemplation” (p. 244). Hobbes views all mental phenomena as reducible to the motions of bodies coming into contact with each other, whether that be tiny corpuscles of light making contact with the retina to explain vision, or whether that be some sort of intra-cranial motion within the nerves of the brain to produce pain. Those that arrive at an understanding of the material causes and effects of motion within the

LYCEUM

mind that produce the various emotions are in a position to understand the motivations of the human will. It is at this point, that Hobbes is able to extend his mechanistic philosophy to his political philosophy.

Hobbes rejects the idea of a *summum bonum* that motivates humanity. Hobbes thinks people will always disagree about what is good, or what is evil, but will always agree on the continuance of what is experienced as pleasurable, and the avoidance of what is experienced as a displeasure. This early version of the Freudian pleasure principle is elucidated quite clearly when Hobbes writes, “and as we call good and evil the things that please and displease, so call we goodness and badness, the qualities or powers whereby they do it” (*EL*, p. 44). Hobbes’ celebrated negative view of human nature makes its appearance in *The Elements of Law* in 1640, though this was unpublished, roughly ten years before Hobbes began work on *Leviathan*. It is clear that Hobbes felt his description of the state of nature is a logical consequence of his conception of human nature and the will, though it is unclear whether or not Hobbes thought the state of nature actually occurred in history. Hobbes imagines humans in their natural state, *sans* any social obligations. Humanity seeks after its desires, and these are most often in conflict with each other. In such a state, war is constant, yet because there is no social obligation, there is no law to be broken. Man has a right to everything, but does not have law. Hobbes writes, “and forasmuch as necessity of nature maketh men to will and desire *bonum sibi*, that which is good for themselves, and to avoid that which is hurtful...it is not against reason that a man doth all he can to preserve his own body and limbs...that which is not against reason, men call right...it is therefore a *right of nature*; that every man may preserve his own life and limbs, with all the power he hath” (*EL*, p. 79).

Many political scholars refer to this *right of nature* as Hobbes’ first principle of political philosophy,⁷ and it does set in motion all that is to follow with regard to the raising of *Leviathan*; it is the motivating factor for the

⁷ See Strauss (1986) and Jaume (2007) for examples of how the *right of nature* principle can be used to argue for Hobbes’ political theory as a source of liberalism—a thesis I wholeheartedly reject.

The Surreptitious Leviathan: Concealing the Beast of Scientific Reason

emergence of the reason that breaks man free of the state of nature and into covenant with one another; however, it is clear from a consideration of Hobbes' geometric method, and our study of Hobbes' theory of knowledge and theory of mind, that this principle is a theorem that is ultimately derived from his theory of motion, whose principles were first elucidated in Hobbes' *Little Treatise*. In short, it follows from Hobbes' earlier thought that a failure to understand Hobbes as a philosopher of motion, even when he discusses social organization and political theory, will inevitably lead to misunderstandings of the various mechanisms at play in Hobbes' political philosophy.⁸ It is precisely this failure on the part of some Hobbes scholars that has led them to identify the Leviathan with the sovereign, or with his political philosophy as a whole.⁹ However, *Leviathan is reason, the reasoning that provides the impetus for man to make covenant and rise out of the state of nature.*

In discussing the three primary motivations that “incline men to peace,” Hobbes identifies fear, desire, and hope which he calls “passions” (*L*, p. 78). Yet Hobbes also includes a rational supplement to these inclinations. He writes, “And thus much for the ill condition which man by mere nature is actually placed in, though with a possibility to come out of it [the state of nature], consisting partly in the passions, *partly in his reason*” (*L*, p. 78).¹⁰ Hobbes then describes reason as that which “suggesteth convenient articles of peace, upon which men may be drawn to agreement” (*L*, p. 78). The Leviathan is the

⁸ Waldman (1974) rejects any psychological egoism in Hobbes' view of human nature in order to view those that seek covenant out of the state of nature as committing sacrificial acts that betray an enlightened self-interest of altruism. Such a reading is only possible if Hobbes' geometric method of analysis and his distinction between the two types of knowledge is de-emphasized in favor of an altruism that is non-existent in Hobbes' descriptions of human nature. Hobbes rejects the possibility of an intrinsic good, so from where would this altruistic form of self-interest be derived?

⁹ Tralau (2007) argues that the use of the Leviathan as a mythological symbol serves a theoretical purpose that contradicts his geometric methodology. His argument identifies the Leviathan with the “otherness” of the “absolute power of the sovereign,” such that there is a contradiction implicit in Hobbes' methodology and his mythological symbolism (77).

¹⁰ I have added the italics here for emphasis.

LYCEUM

rationality that makes possible the second form of knowledge; the knowledge that separates us from the beasts and our beast-like existence when man is observed in the state of nature; the knowledge of science—obtained through the very method Hobbes discovered when he first uncovered Euclid in Geneva in 1628. Again in the *Leviathan*, Hobbes writes, “The other [form of knowledge] is of consequences, and is called science, the record of which is usually called philosophy” (p. 49).

It is the *Leviathan* that is the *lex naturalis* that compels man to divest his natural right in favor of the law of the sovereign. Mankind in the state of nature uses his reason to establish his right, but this type of reason cannot establish law. It is evident that there are two types of reason at play in Hobbes, just as there are two types of knowledge and two types of natural law. The type of reason that yields natural right is the type of reason that the beasts employ when they develop prudential knowledge. This is a type of instrumental reasoning that involves a mere extrapolation from past experiences—one can imagine a human being in the state of nature reasoning as follows: “The last person I encountered stole my food, thus I should either kill the next person I meet, so that I won’t have my food stolen, or I should hide my food and pretend to be without food to steal.” Hobbes writes of a different kind of reason when he claims, “there can be no other law of nature than reason, nor no other precepts of NATURAL LAW, than those which declare unto us the ways of peace, where the same may be obtained” (*EL*, p. 82). This is a law that is issued by human rationality, as distinct from the reasoning powers of the beasts, and as distinct from a law that is issued by divine command. Again Hobbes appeals to what I am calling the *Leviathan*, when he writes, “reason therefore dictateth to every man for his own good, to seek after peace, as far forth as there is hope to attain the same” (*EL*, p. 81). In the *Leviathan*, Hobbes expands upon this point by describing two laws of nature. The first, which he calls “fundamental,” is *contained* within man’s right of nature such that “*he may seek and use all helps and advantages of war*” only whence he cannot obtain the peace “that every man ought to endeavor” (p. 80). From this more fundamental law of nature, Hobbes derives the second law of

The Surreptitious Leviathan: Concealing the Beast of Scientific Reason

nature which provides for a “contract in way of peace”—a covenant through which man lays down his right of nature (*L*, p. 80).

If the unity thesis is defensible, then Hobbes’ *Leviathan* must be understood in reference to his mechanistic materialism. This is quite problematic when one considers the manner in which Hobbes’ opens the book:

For seeing life is but a motion of limbs, the beginning whereof is in some principal part within, why may we not say that all *automata*...have an artificial life? For what is the *heart*, but a *spring*; and the *nerves*, but so many *strings*; and the *joints*, but so many *wheels*, giving motion to the whole body, such as was intended by the artificer?
(*L*, p. 3)

This mechanical description of man’s anatomy mirrors the mechanical description of Hobbes’ Commonwealth, and is consistent with Hobbes’ mechanistic materialism as described in *De Corpore* and *The Elements of Law*. However, Hobbes employs this analogical symbolism throughout his brief introduction and the use of mythological imagery and rhetorical flair in *Leviathan* marks a distinct change of style from his earlier writings. This change in style does not indicate a change in his methodology. Hobbes’ method in *Leviathan* remains strictly geometric. The first 16 chapters can be seen as summaries of his work in *De Corpore* and *The Elements of Law*; Hobbes briskly runs through his thoughts on his mechanistic materialism, his philosophy of language, his epistemology, and his distinction between the two versions of natural law and the right of nature. This summary exposition of his thought can, indeed, lead one to think that the method of deduction from first principles employed in *De Corpore* and *The Elements of Law* has been abandoned in *Leviathan*; however, Hobbes himself clarifies the issue when he writes in reference to his (as then unpublished) work in *De Corpore* and *The Elements of*

LYCEUM

Law, “I have elsewhere written of the same [the natural causes of sensory experience] at large. Nevertheless, to fill each part of my present method, I will briefly deliver the same in this place” (*L*, p. 6). It is clear that Hobbes feels that his methodology *requires* him to discuss his mechanistic materialism; however, it is also clear that Hobbes feels he has bigger fish to fry.

Hobbes writes *Leviathan* in Paris, while England is engaged in a bitter Civil War. King Charles I has been executed and the other English Royalists in Paris are unsure of what to do—ought one support the Prince of Wales and continue the fight, or ought one accept the new Commonwealth (established in 1650, a year prior to the completion of *Leviathan*) and the ‘terms of engagement’ that were offered for their return?¹¹ Hobbes addresses the diehard Royalists when he writes in the ‘Review and Conclusion’ to *Leviathan*, “and though in the revolution of states there can be no good constellation for truths of this nature to be born under...I cannot think it [the new Commonwealth] will be condemned at this time, either by public judge of doctrine, or by any that desires the continuance of public peace” (p. 497). Hobbes also addresses the new Commonwealth when he warns, “I have set down for one of the causes of the dissolutions of commonwealths their imperfect generation, consisting in the want of an absolute and arbitrary legislative power” (*L*, p. 491). Like Machiavelli, Hobbes envisions his *Leviathan* as a textbook in political philosophy that can prevent future civil wars if the King of the new Commonwealth follows its mandates, but he also views it as clarion call to the defeated Royalists urging them to finally come home. Hobbes crafts *Leviathan* for a much broader audience than his previous works, and such considerations help explain Hobbes’ shift in style and his de-emphasis upon the methodology that is vital to his conception of the proper manner of philosophizing. *Leviathan* needed to be a book that pulled at the heartstrings as well as one that appealed to scientific

¹¹ The Terms of Engagement stipulated that many of the Royalists in Paris could return to England if they paid a heavy fine and swore an Oath of Engagement to the New Commonwealth, and their titles and lands would be returned to them. See Martinich (1999, pp. 219-25) for a detailed description of “the Engagement Controversy.”

The Surreptitious Leviathan: Concealing the Beast of Scientific Reason

reasoning. It is for this reason that he likens the creation of the Commonwealth to an art that “created that great LEVIATHAN called a COMMONWEALTH...which is but an artificial man” (*L*, p. 3). It is best to read Hobbes’ introduction to *Leviathan* as a work of rhetorical irony, what Hobbes’ literally says—the claim that the Leviathan *is* the Commonwealth—should be understood as an intentional oversimplification, written in order to help reinforce the perception that the head of such a Commonwealth, the sovereign, is absolutely powerful. In order to discover what Hobbes literally means, one would need to consider his entire philosophy in both its geometric methodology and its mechanistic roots, to discover that the birth of the Leviathan is an event in which human reason distinguishes itself from the madness of the beast-like state of nature, where the Commonwealth is understood as the product of Leviathan.¹²

It is at this point that we will venture away from Hobbes’ political philosophy as contained within the pages of *Leviathan*. We have established a congruence between the methodology involved in *De Corpore* and *The Elements of Law* and the methodology employed, though downplayed, in *Leviathan* with respect to the genesis of Hobbes’ political philosophy. We have shown, through careful study of Hobbes’ *De Corpore* and *The Elements of Law*, the manner in which Hobbes envisions his political thought as derivable from his work in natural philosophy, particularly his conception of human nature as being reducible to the laws of motion. Furthermore, we have placed *Leviathan* within its proper historical context to explain the manner in which Hobbes’ title monster is best understood as representative of Hobbes’ conception of the reason that yields both scientific knowledge and political science; after all, Hobbes did not

¹² Though it is beyond the scope of this paper, I believe that this interpretation of what the Leviathan actually represents can shed new light onto the debate amongst Hobbes scholars regarding his strange manner of scriptural interpretation, his atheism, as well as his negative view of human nature *sans* Leviathan, i.e. humanity devoid of scientific reasoning, as in the divine command theory accepted prior to *Leviathan*. All of which would be better understood (and more acceptable interpretations), if viewed as the consequence of his mechanistic philosophy and his geometric methodology.

LYCEUM

feel *Leviathan* was a work of political philosophy *qua* political philosophy *circa* 1651, but Hobbes felt *Leviathan* was a work that established the first political *science*.

Yet, if we are to understand the underlying motivation for viewing Hobbes' *Leviathan* as a surreptitious work, we must consider the reaction to the book from Hobbes' contemporaries. Hobbes was no doubt aware of the sort of controversy the book would generate.¹³ Armed with these considerations of what sort of firestorm his 'little pamphlet' might generate, Hobbes wrote *Leviathan* with the sort of surreptitious irony he thought necessary to pull the wool over the eyes of those that would not understand the true meaning of the text, while pacifying those that were prepared to accept the consequences of the new science. Unfortunately, or perhaps fortunately, for the sake of his neck, Hobbes failed in this effort. Too many of his contemporaries rejected his mechanistic materialism, balked at his conception of human nature, and demanded retribution for his blasphemous scriptural interpretations, while a scant few understood that *Leviathan* was meant to be a precise description of man's new God in action. Those that understood this were prepared to follow the path of the Leviathan through to Hobbes' conclusions, however frightening these conclusions might be. Those that failed to understand the Leviathan understood only enough to reject Hobbes based upon his various 'theorems' and 'secondary principles' without understanding from where these principles were derived. It is for this reason that we will understand *Leviathan* as a work of surreptitious irony that conceals the Leviathan of scientific reason beneath its symbolism and rhetoric.

¹³ Though Hobbes' belief that the book would force Aristotelian Scholasticism from the schools of England reflected a rather naïve faith, not in scientific reason, but in humanity's readiness to accept this new type of reason, Hobbes still understood that the book would face serious opposition from his contemporaries.

The Surreptitious Leviathan: Concealing the Beast of Scientific Reason

The Negative Reception of *Leviathan*

Hobbes' denial of the *summum bonum* was a controversial point of disagreement between Hobbes and his contemporaries. Hobbes' claim that there is no moral good or evil that is not a relative good or evil and that man is motivated strictly by self-interest, although latter echoed in the 18th Century by Bernard Mandeville, places him strictly in the minority with respect to English Moralists from the 17th century and 18th Century.¹⁴ G. A. J. Rogers (2007), commenting on the English reception of Hobbes' moral theory, writes "his philosophy was seen as implying a subjectivism about moral theory that opened the floodgates to wantonness and vice" (p. 425). Daniel Scargill, a former Hobessian forced to publish a recantation of Hobbes in 1669, describes Hobbes' moral theory as "founded only in the law of the civil magistrate" (as cited in Parkin, 2007b, p. 448). This damaging admission helped foster public opinion of *Leviathan* as a book that would lead one into moral corruption. Thus, the British Moral Theorists of the 18th century often chose Hobbes as their foil in their arguments for their respective moral theories. Seth Ward, in 1651, felt Hobbes' fatal error was in Hobbes' refusal to recognize that "judgments...about good and evil are made at the ordaining, and almost incitement, of nature, and before Hobbes all men were accustomed also to call them justice and injustice" (as cited in Parkin, 2007a, p. 167). Thus, Ward feels Hobbes has reduced moral law to the whim of the sovereign, as opposed to a law of nature. Ward's interpretation is half correct. Hobbes' moral theory does indeed depend upon the law of the sovereign; however this law is indeed the law of the Leviathan—the scientific reasoning that produces scientific law, which is further reducible to Hobbes'

¹⁴ For an entertaining version of Hobbes' rejection of *summum bonum*, described within English society as opposed to the state of nature, please see Mandeville's (1924) *The Fable of the Bees*, originally published in various installments in the years between 1720 and 1730. Contrast this with the dominance of the moral sense theorists of 18th century England (Shaftesbury, Hutcheson, Butler, Balguy, etc.).

LYCEUM

theory of motion. Again, Hobbes' *Leviathan* proves surreptitious in that it does not explicitly emphasize the connection between his moral theory and his mechanistic materialism.

While subsequent political theorists adopted many of Hobbes' ideas in their political philosophy, they often went to great lengths to either disguise or deny the influence Hobbes' political philosophy had upon their work. Others would borrow Hobbes' political ideas while "nest[ing] them in theoretical frameworks at odds with Hobbes' theory" (Parkin, 2007a, p. 138). Thomas White's political theory rejects Hobbes' emphasis upon fear with regard to the motivating factors involved in achieving rational consensus within the state of nature.¹⁵ White's political theory borrows from Hobbes when considering the need for covenant under a sovereign power, but White's motivation is Aristotelian, in that the desire for covenant is not a fearfully motivated self-interest on the part of those under the state of nature, but a recall to man's natural sociability. Man's social nature is here understood to be paramount, in opposition to Hobbes' right of nature which White believes to be "against all generosity, and embraces Nature it self...agreeth neither with Philosophy nor Morality" (as cited in Parkin, 2007a, p. 47). Unlike Hobbes, White rejects man's instrumental reasoning in the state of nature as a mere embracing of nature, while White holds to a form of reasoning that recognizes love and social desire as that which pulls man from his base state. This difference from Hobbes is slight but substantial. Hobbes' instrumental reasoning on the part of man in the state of nature must be distinguished from the scientific reasoning involved in man's decision to make covenant, if Hobbes' theory of knowledge is to retain his distinction between prudential knowledge and scientific knowledge. Furthermore, unlike White, Hobbes views self-interest as the motivating factor for the occurrence of both types of reasoning, and this must be the case, otherwise an inherent sociability within man would imply scientific reasoning as being inherent in man, yet *not* inherent in nature, a thesis that would contradict Hobbes' mechanistic materialism with respect to human nature.

¹⁵ See White (1655).

The Surreptitious Leviathan: Concealing the Beast of Scientific Reason

John Hall's 1654 work, *Of Government and Obedience*, defends absolute rule by a sovereign power as well; however, Hall denies Hobbes' conception of the state of nature as a state of war because Hall believes such a conception of the state of nature implies that "they [men in the state of nature] could be left by a careful God in such a confused condition, where...they should fall to slaughter of one another, till their bleeding wounds, and not His Precepts of Providence, had taught them rules of Subjection" (as cited in Parkin, 2007a, p. 79). Hall's argument suggests that an *omni-benevolent* God cannot permit a state of nature that is warlike, such that fear of war prompts man to covenant, rather than fear of God. Hobbes' political theory rejects a divine right model in which the sovereign becomes sovereign due to the will of God—Hobbes' model does not require the providence of God, while Hall's model does require the providence of God. Under Hobbes' model, individuals within the state of nature agree to covenant under the sovereign as protectorate and representative of "all their power and strength upon one man, or upon one assembly of men, that may reduce all their wills, by plurality of voices, unto one will, which is as much as to say, to appoint one man or assembly of men to bear their person, and every one to own and acknowledge himself to be author of whatsoever he that beareth their person shall act" (*L*, p. 109). The authority of the sovereign is not derived from, and is thus independent of, the authority of God. Hobbes' cause of the generation of the commonwealth is the individual's giving up of his rights in the state of nature to the law of the sovereign, which, at least in Hobbes' optimism, is co-extensive with the laws of nature, unless the sovereign no longer has "the power...by which he is able to protect them [his subjects]" (*L*, p. 144). Hobbes implicitly argues against the divine right model when he writes, "a subject that has no certain and assured revelation particularly to himself concerning the will of God is to obey for such the command of the commonwealth" (*L*, p. 188). This passage can indeed be interpreted as leaving the door open for divine command if it is considered separately from Hobbes' later scriptural interpretations; however, Hobbes makes it clear that the age of prophecy is over when he writes

LYCEUM

(after considering several scriptural descriptions of prophecy that involve visions and dreams), “for when Christian men take not their own Christian sovereign for God’s prophet, they must take their own dreams for prophecy they mean to be governed by, and the tumor [tumult] of their own hearts for the Spirit of God... and by this means destroy all law, both divine and human, reduce all order, government, and society to the first chaos of violence and civil war” (*L*, p. 293). Hobbes’ warning is clear. Those that *think* they are following divine command are incorrect, because divine command cannot contradict the will of the sovereign, and this mistaken prophecy can lead to civil war, or a return to the state of nature. Though Hall obviously borrows some of Hobbes’ political ideas, he employs them to different ends—the ends of identifying the sovereign as subject to the will of God.

Of all the points of disagreement between Hobbes and his contemporaries, Hobbes’ perceived atheism was the most damaging and threatening accusation hurled at Hobbes in response to *Leviathan*. Whether or not Hobbes was an atheist is not pertinent to our present considerations.¹⁶ It is pertinent that a common attack of Hobbes’ philosophy by his contemporaries centered on his manner of interpreting scripture to support his political philosophy, no matter how unorthodox the interpretation. Due mainly to these unorthodox interpretations, Hobbes was branded an atheist by many writers, yet this brand was not often a public brand. As Jon Parkin points out, “Hobbes’s early critics tended to shy away from public accusations of atheism, and what is remarkable is not the volume of such accusations, but rather how few there were, and how cautiously they were made” (Parkin, 2007a, 133). Early criticisms of Hobbes’ theology focused upon the heretical nature of his scriptural interpretation, not on any claim that their author was an atheist. Alexander Ross compared Hobbes’

¹⁶ For detailed cases supporting the claim that Hobbes was an atheist, see Jesseph (2002), Curley (1992), or Skinner (1996). For the opposite position, see Taylor (1965), Warrender (1957), or Zarka (1996).

The Surreptitious Leviathan: Concealing the Beast of Scientific Reason

scriptural interpretations to those of “Cerinthus the heretick.”¹⁷ When Hobbes writes of the resurrection and the Kingdom of Christ as an *earthly* kingdom, Hobbes unconventional scriptural interpretation allows his contemporaries to attack his theology as heretical.¹⁸ Yet, in private letters, Hobbes’ name was besmirched by the brand of atheism, to the public effect of Hobbes’ removal from the English court in Paris; Sir Edward Nicholas expresses this private opinion in the following:

I hear Lord Percy is much concerned in the forbidding Hobbes to come to Court, and says it was you and other Episcopal men, that were the Cause of it. But I hear that Wat. Montagu and other Papists (to the shame of the true Protestants) were the chief Cause, that the grand Atheist was sent away. (As cited in *Correspondence*, 1994b, p. 800)

This sentiment would slowly grow until Hobbes’ name was publicly linked to atheism. Thomas Tenison writes, “of late, [Hobbes] hath set forth his *Leviathan* in the Latine Tongue; declaring his desire of spreading his Malady throughout the World” (as cited in Parkin, 1997a, p. 174). In fact, Hobbes’ name would soon become a negative epitaph for any poorly received work. Parkin notes, “after contributing to Hobbes’s pamphlet he [Henry Stubbe] was accused of Hobbism by the Presbyterian Daniel Cawdry” (Parkin, 2007a, p. 174). George Lawson attacks Hobbes’ scriptural interpretation as written in “such a loose and impious abusive manner, that I verily perswade my self, he doth not believe them

¹⁷ See Ross (1653, pp. 35-6). Parkin describes in footnote 114 Cerinthus’ distinction between Jesus and Christ, where Jesus “was a mere man, though eminent in holiness. He suffered and died and was raised from the dead...and will be raised from the dead at the Last Day and all men will rise with him” (Parkin, 2007, p. 125).

¹⁸ See *Leviathan*, (p. 302, [3]).

LYCEUM

to be revealed and written from heaven, or that Jesus Christ was an ordinary just man, much less the Eternal Son of God Incarnate” (as cited in Parkin, 2007a, p. 50).

It is my contention that much, if not all, of Hobbes’ scriptural interpretations are meant in a surreptitiously ironic manner. They are not to be taken literally, but are meant to cover over Hobbes’ true rationale for presenting them—they are the theological consequences of his political philosophy, which in turn, is a consequence of his mechanistic materialism. Hobbes ought always to be thought of as a philosopher of motion first, a political philosopher, moral theorist, and theologian, *after* this fact. Henry Rosenthal (1989) strikes this chord with a strangely Heideggerian description of Hobbes when he describes *Leviathan* thusly, “the flourish of this title is truly esoteric; it conceals by revealing” (p. 32). Hobbes, in revealing his political philosophy and its theological implications in *Leviathan* conceals the very mechanistic materialism that grounds his thought. Doug Jesseph (2002), in consideration of the manner in which Hobbes ironically interprets scripture, writes, “Hobbes himself approached problematic passages in the Scriptures with precisely the kind of critical “sifting out of Philosophicall truth by Logick” that he claimed to be inappropriate” (p. 157). Edwin Curley (1992) suggests that Hobbes’ philosophy in *Leviathan* should be understood as intentionally less bold than it could have been; Curley thinks Hobbes employs a “suggestion by disavowal” approach, in which the author leads one to a conclusion, before disavowing the conclusion just as it begins to seem logically consequent—one can imagine Hobbes defending himself in such a case with the phrase, “you said it, not I!”

The majority of Hobbes’ 16th Century contemporaries who attacked him so viciously on his moral theory, his political theory, and his theological interpretations, failed to attack Hobbes at his most critical point—his mechanistic conception of the universe and the scientific reasoning that is the true *Leviathan*. Henry More and Ralph Cudworth, two Cambridge Platonists, were two of Hobbes’ contemporaries that *did* see through the disguises of

The Surreptitious Leviathan: Concealing the Beast of Scientific Reason

Hobbes' *Leviathan* to the mechanistic materialism at its core. Rogers (2007) describes their main disagreement with Hobbes' philosophy in the following:

The crucial premise for both of them is that materialism is a philosophy of both mind and nature that is quite incapable of doing justice to the facts of human experience. More had in his early studies come to a great admiration of the philosophy of Plotinus as well as Plato, and it was the undeniability of a spiritual agent or force that could not be squared with [the] materialism that dominated philosophy. (p. 431)

More thought that Hobbes' mechanistic materialism required a material God, yet God could only exist as a spiritually immaterial being, therefore Hobbes' mechanism, and with it, Hobbes' political philosophy, moral philosophy, and theology, ought also to be rejected. Cudworth was more direct in his attack on Hobbes but for similar reasons as More. Cudworth writes, "that prodigious paradox of Atheists, that cogitation itself is nothing but local motion...a modern atheistic pretender, to wit, hath publicly owned his name to this same conclusion, that 'mind is nothing else but local motion in the organic part of man's body'" (as cited in Rogers, 2007, p. 432). Cudworth argues that a materialist philosophy fails to adequately describe human nature. Both of these rationalist thinkers saw Hobbes' *Leviathan* as dependent upon a mechanistic materialism that was unacceptable. They were one of the few, though they were *not* influential, English contemporaries of Hobbes whose rejection of Hobbes' political philosophy and theology, was a consequence of their rejection of his underlying mechanistic theory of motion.

LYCEUM

A Surreptitiously Ironic Reading of Leviathan

We have explicated the degree to which the political philosophy and the scriptural interpretations described in Hobbes' *Leviathan* can be read as a consequence of his mechanistic materialism, and more particularly, his first principles reducing sensation to bodies in motion, and only bodies in motion. It is not my intention to demonstrate the truth of the claim that all of Hobbes' philosophy necessarily follows from his mechanistic conception of the universe; we only need to understand Hobbes' *Leviathan* as Hobbes understood it—the logical consequence of a mechanistic universe. This scientific reasoning, this logic that Hobbes' believes all things are in deference to, is indeed his mortal God, the Leviathan. Thus, we now understand the book *Leviathan* as a description of *the* Leviathan—a description of the event in which an artificial man, man in the state of nature, creates an artificial God, the commonwealth, in the image of his mortal God, Leviathan.

Man is assumed to exist in the state of nature prior to this monumental event. This man, very much like the beasts, is motivated strictly by self-interest and employs instrumental reasoning to wage war on his neighbors. And then, the event occurs; a thinker, motivated by the scientific reasoning of the new science, applies Euclid to human nature. The Leviathan is called up from the depths and emerges in the form of the commonwealth; a commonwealth created by trust, a trust that transcends the self-interestedness that creates it. Man's initial state is transcended and the possibility is born for man to govern himself, instead of being governed by the myth of his youth. The old mysteries are explained by the new science; the world is disenchanted. Whether or not God exists is a question no longer fundamental to man—in its place, a new question arises; is man prepared to accept the consequences of the Leviathan?

Can man accept this new, will-less existence? For in the commonwealth, man has given up his right of nature, in favor of the law of nature. His will is usurped by the will of the sovereign—a figurehead for the law of Leviathan. Can man accept an ex-istence in which God cannot be appealed to as the ultimate

The Surreptitious Leviathan: Concealing the Beast of Scientific Reason

arbiter in difficult disputes? It is my claim that Hobbes' *Leviathan* is surreptitiously ironic, because Hobbes' feared that his society would not accept a will-less existence, or a God-less existence, yet the historical milieu in which Hobbes found himself, required that he *finally* publish his mature thought. The new commonwealth must be educated as to the proper manner of governance, if it is not to slip into another civil war. The Royalist expatriates must give up the old fight—they, and Hobbes, had lost. Hobbes was faced with the daunting task of presenting a coherent political philosophy that was a consequence of the methodology and mechanism which underwrites his thought, while at the same time, concealing his mechanistic materialism so as not to have England throw out the baby with the bath water. *Leviathan* fulfilled this task, while at the same time branding its author an enemy. In 1657, a Hobbes who seems tired and worn down by his critics, writes to his friend, Samuel Sorbière, "you will become a perpetual object of hatred to the doctors, just as I am (because of my political theory) to the theologians. The kingdom of truth is not of this world, but the next" (1994b, p. 448).

Had the Leviathan become Hobbes' failed God? One imagines Hobbes running the English countryside in pursuit of his creation, like Dr. Frankenstein, willing to destroy the beast through recantations and apologies. Rogers (2007) writes, "His picture of knowledge was a unified one, and he may well have believed that if he were wrong on this front [mathematics]...then his whole philosophy would have collapsed" (p. 435). Jesseph (1999) notes that Hobbes had committed the same error in mathematics—his claim to have squared the circle—which he identified in *Leviathan*. Hobbes writes, "and as in Arithmatique, unpractised men must, and Professors themselves may often erre, and cast up false; so also in any other subject of Reasoning, the ablest, most attentive, and most practised men, may deceive themselves, and inferre false Conclusions" (as cited in Jesseph, 1999, pp. 355-6). Perhaps Hobbes' failure in mathematics, colored his opinions of his entire philosophy, then again, Hobbes' fought his battles in mathematics to the end of his life. As A. P. Martinich (1999)

LYCEUM

writes, “His aches and pains notwithstanding, Hobbes was not going to “go gentle in that good night;” He published three large-scale geometric works in the last decade of his life” (p. 336).

Our surreptitiously ironic reading of *Leviathan* remains. John Dewey (1974) writes of Hobbes’ conception of his philosophy:

Whether or not he believed in a theological God, he did believe that reasoning was divine, and that there is a sincere piety toward reason in his regarding rational precepts as divine; and that accordingly he believed in some genuine sense that God was reason. (p. 27)

Long after its creator has passed, the Leviathan still lives. Like Hobbes’ 17th century contemporaries, we are frightened of its power. Unlike Hobbes, we are unwilling, or unable to view ourselves in the light of our creation—this artificial God may indeed reflect ourselves, yet the truth grinding in its internal gears is a truth we are not prepared to accept. The Leviathan remains surreptitiously underwater; we cannot accept its monstrous logic.

References

- Brandt, F. (1927). *Thomas Hobbes’ mechanical conception of nature* (V. Maxwell and A. Fausbøll, Trans.). Copenhagen: Levin and Munksgaard Press.
- Curley, E. M. (1992). Religion and morality in Hobbes. In Jules L. Coleman and Christopher W. Morris (Eds.), *Rational commitment and social justice: Essays for Gregory Kavka* (pp. 90-111). Cambridge: Cambridge University Press.

***The Surreptitious Leviathan:
Concealing the Beast of Scientific Reason***

- Curley, E. M. (1998). 'I durst now write so boldly' or, how to read Hobbes's theological-political treatise. In Daniela Bostrengi (Ed.), *Hobbes e Spinoza, Atti del Convegno Internazionale, Urbino 14-17 ottobre, 1988* (pp. 497-593). Naples: Bibliopolis.
- Dewey, J. (1974). The motivation of Hobbes's political philosophy. In R. Ross, H.W. Schneider, & T. Waldman (Eds.), *Thomas Hobbes in his time* (pp. 61-83). Minneapolis: University of Minnesota Press.
- Hobbes, T. (1994a). *Leviathan: with selected variants from the Latin edition of 1668* (E. Curley, Ed.). Indianapolis: Hackett Publishing Company.
- Hobbes, T. (1994b). *The correspondence of Thomas Hobbes* (Noel Malcolm, Ed.). In *The Clarendon Edition of the Works of Thomas Hobbes* (Vols. 6-7). Oxford: Oxford University Press.
- Hobbes, T. (1999). *Oxford World Classics: Thomas Hobbes' The Elements of Law and De Corpore Politico w/ appended chapters from De Corpore* (J. C. A. Gaskin, Ed.). Oxford: Oxford University Press.
- Jaume, L. (2007). Hobbes and the philosophical sources of liberalism. In P. Springborg (Ed.), *The Cambridge Companion to Hobbes's Leviathan* (pp. 199-216). Cambridge: Cambridge University Press.
- Jesseph, D. M. (1999). *Squaring the circle: The war between Hobbes and Wallis*. Chicago: University of Chicago Press.
- Jesseph, D. M. (2002). Hobbes's atheism. In A. French & H. Wettstein (Eds.), *Midwest Studies in Philosophy*, 26(1), 140-66.

LYCEUM

- Leijenhorst, C. (2007). Sense and nonsense about sense: Hobbes and the Aristotelians on sense perception and imagination.” In Patricia Springborg, *The Cambridge Companion to Hobbes’s Leviathan* (pp. 82-108). Cambridge: Cambridge University Press.
- Mandeville, B. (1924). *The fable of the bees, or private vice, publick benefits* (Vols. 1-2).. Oxford: Clarendon Press.
- Martinich, A. P. (1999). *Hobbes: A biography*. Cambridge: Cambridge University Press.
- Mintz, S. I. (1962). *The hunting of Leviathan: 17th century reactions to the materialism and moral philosophy of Thomas Hobbes*. Cambridge: Cambridge University Press.
- Parkin, J. (2007a). *Taming the Leviathan: The reception of the political and religious ideas of Thomas Hobbes in England 1640-1700*. Cambridge and New York: Cambridge University Press.
- Parkin, J. (2007b). The reception of Hobbes’s *Leviathan*. In Patricia Springborg (Ed.), *The Cambridge Companion to Hobbes’s Leviathan* (pp. 441-59). Cambridge: Cambridge University Press.
- Rogers, G. A. J. (2007). Hobbes and his contemporaries. In Patricia Springborg (Ed.), *The Cambridge Companion to Hobbes’s Leviathan* (pp. 413-40). Cambridge: Cambridge University Press.
- Rosenthal, H. M. (1989). *The consolations of philosophy: Hobbes’s secret; Spinoza’s way* (A. L. Rosenthal, Ed.). Philadelphia: Temple University Press.

***The Surreptitious Leviathan:
Concealing the Beast of Scientific Reason***

- Skinner, Q. (1996). *Reason and rhetoric in the philosophy of Thomas Hobbes*. Cambridge: Cambridge University Press.
- Strauss, L. (1986). *Droit naturel et histoire*. Paris: Flammarion Press.
- Taylor, A. E. (1965). The ethical doctrine of Thomas Hobbes. In K. Brown (Ed.), *Hobbes Studies* (pp. 35-56). Oxford: Oxford University Press.
- Tralau, J. (2007). Leviathan, the beast of myth: Medusa, Dionysos, and the riddle of Hobbes's sovereign monster. In Patricia Springborg (Ed.), *The Cambridge Companion to Hobbes's Leviathan* (pp. 61-81). Cambridge: Cambridge University Press.
- Waldman, T. (1974). Hobbes on the generation of a public person. In R. Ross, H.W. Schneider, and T. Waldman (Eds.), *Thomas Hobbes in his time* (pp. 61-83). Minneapolis: University of Minnesota Press.
- Warrender, H. (1957). *The political philosophy of Hobbes: His theory of obligation*. Oxford: Oxford University Press.
- Zarka, Y.-C. (1996). First philosophy and the foundation of knowledge. In T. Sorell (Ed.), *The Cambridge Companion to Hobbes* (pp. 62-85). Cambridge: Cambridge University Press.

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

Leonard Ferry

Near the middle of the twentieth century one finds a widespread consensus among Catholic political philosophers and theologians with respect to the Church's teaching on the means through which the *right* to govern is acquired by political authority. There are two principles at the center of this consensus. First, God is taken to be the source of all authority, including the power to bind the conscience—that is, the power to impose obligations by precept and command that are normative for human beings. Second, political authority specifically is understood to be given by God to the people as a whole, and the people are held to transfer the same to the respective political offices or institutions, though in rare cases of direct democracy they retain and exercise this power themselves. When I write of the *consensus view*, therefore, I will be referring to the two principles taken together as an account of the origin of and justification for the authority of rulers and the correlative obligations of the ruled.¹ In *Scholasticism and Politics* Jacques Maritain (1940) identifies the consensus view's first principle as the 'Pauline principle': "This idea in the Christian tradition is classically expressed by the Pauline principle that all authority derives from God as from its primordial source" (p. 104). The second principle of the consensus view was generally referred to as the *transmission theory*. The two principles are inseparable for supporters of the consensus view, even though it is more common to find references only to the latter—that is, the transmission principle or the *transmission theory*. Maritain's reference to the first principle—the

¹ The principles are glanced at in Antony Black's *Political Thought in Europe 1250-1450* (1992). Black mentions both principles without identifying them as such or explaining their inter-relatedness: "Kings derived their authority from God...And kings were at the same time said to derive their authority from the people—the political community, however structured. This was a mysterious as much as a constitutional belief: as the constitutional monarchist Sir John Fortescue put it, 'from the people there breaks forth a kingdom, which is a mystic body governed by one man as head'" (137). Twentieth century defenders of the consensus view do not restrict its application to constitutional or other types of monarchy, but the core ideas remain the same.

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

Pauline principle—is a helpful reminder, however, that the consensus view consists of two principles *and* that the second principle is dependent on the first. Absent the origin of authority in God, there would be no authority in the people to be transferred, and the fact that political power regularly binds individuals to perform or refrain from performing specific acts would be based solely on its coercive force. Put simply, political authority would be a matter of might rather than right.

The consensus view was widely held in the first half of the twentieth century among Catholic philosophers and theologians, and the figures who supported some version of it constitute an impressive list of mid-century Thomists, including Heinrich Rommen (1945), Wilfrid Parsons (1939, 1940a, 1940b, 1941),² Yves R. Simon (1940, 1966, 1968, 1980), Charles N. R. McCoy(1989), and, of course, Jacques Maritain. Strikingly, today many of those who are interested in the political philosophy of Aquinas and the natural law tradition more generally are silent about the consensus view.³ Mark Murphy

² See also “St. Thomas Aquinas and Popular Sovereignty,” *Thought: Fordham University Quarterly* 16 (September 1941): 473-92. In this article Parsons helpfully reminds us that democratic government is but one species and cannot be substituted for the genus of good—that is, legitimate—forms of government. Legitimacy requires “a rule that is for the common good, is representative of the people, and is derived for the ruler immediately from the community itself” (474). Democracy adds to these three elements of the genus of good government the manner or mode of exercise—rule by the people—as a specific difference. It is a mistake to confuse democracy with the legitimacy of political authority.

³ Robert Pasnau and Christopher Shields usefully describe the “eclectic” nature of Aquinas’ ethical theory—“It is, all at once, a virtue theory and a natural law theory, with divine commands playing a role as well. It combines deontological and consequentialist aspects and in addition has a strong teleological component”—but one might read their introduction to the thought of Aquinas without realizing that he had anything whatsoever to say about politics (2004, p. 217). Much the same can be said of Rebecca Konyndyk DeYoung, Colleen McCluskey, and Christina Van Dyke’s *Aquinas’s Ethics/Metaphysical Foundations, Moral Theory, and Theological Context* (2009). Paul E. Sigmund’s “Law and Politics” survey essay manages to link Aquinas’ political theory to that of Aristotle, but is silent on the consensus view [(1993, pp. 217-31); the essay was reprinted more recently in *Thomas Aquinas: Contemporary Philosophical Perspectives*, ed. Brian Davies (Oxford

LYCEUM

(2009), for example, in *Natural Law in Jurisprudence and Politics*, makes no reference to either of its principles, despite offering what he describes as a natural law defence of political authority that is indebted to the work of Aquinas. Similarly, no mention is made of the principle in D. E. Luscombe's (1982) survey article, "The state of nature and the origin of the state" in *The Cambridge History of Later Medieval Philosophy*. In Jean Porter's (2010) *Ministers of the Law*, what she calls the "Pauline view" is described as "inadequate" because "it was difficult or impossible on this account to generate criteria on the basis of which to distinguish between legitimate and illegitimate uses of authoritative power" (pp. 46-7). Unfortunately, her version of the Pauline view seems deeply confused. She characterizes it thus: "At the beginning of the scholastic period, many Christians, scholars and lay men and women alike, would have affirmed the Pauline view that human authority rests on a kind of divine authorization, conferred for the purpose of restraining the effects of sin" (p. 47). Insofar as the Pauline principle identifies God as the source of the power to bind the consciences of others, it seems unnecessary to speak of a "kind of divine authorization," and the confusion of the Pauline principle with Augustine's view

University Press, 2002), pp. 325-37]. Michael Baur's "Law and Natural Law" at least touches on the Aristotelian naturalism of Aquinas' theory, specifically the close connection between authority and the common good, but remains silent on political authority's foundation in Paul's letter [in *The Oxford Handbook of Aquinas*, eds. Brian Davies and Eleonore Stump (2012, pp. 238-54). Thomas S. Hibbs takes pains to reduce any potential role of authority by tracing the normativity or obligatoriness of precepts to their intelligibility, rather than to their promulgation by authority [*Virtue's Splendor/Wisdom, Prudence, and the Human Good* (2001, p. 73). Of course, Hibbs is correct to emphasize the intelligibility of normative prescriptions, but between two equally intelligible commands (say, drive on the right or the left-hand side of the road) what determines the law for a given community must be the authority responsible for that community. How does a person or institution come to hold the right, in such cases, to make such decisive determinations? That is what is at issue for Aquinas in tracing the origins of political authority back to the Creator of human nature, a nature that is understandable in part through the normal operation of human reason—that is, the use of reason unaided by revelation.

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

that political authority is exercised for the restraining of human sinfulness only serves to obscure matters more.⁴

In stark contrast to those who simply neglect the Pauline and transmission principles, John Finnis (1980) has openly rejected the latter:

Consent, transmission, contract, custom—none of these is needed to constitute the state of affairs which (presumptively) justifies someone in claiming and others in acknowledging his authority to settle co-ordination problems for a whole community by creating authoritative rules or issuing authoritative orders and determinations. (pp. 248-9)

Transmission is only one in a list of attempted justifications of political authority that Finnis dismisses, but it attracts, as other members in the list do not, a direct refutation from Finnis.⁵ Part of his objection amounts to his belief that the

⁴ Augustine's views are clearly summed up in the following: "the state...is an external order; the peace that it maintains is external peace—the absence, or at least the diminution, of overt violence. The state is also a coercive order, maintained by the use of force and relying on the fear of pain as its major sanction for compliance to its commands. It has no weapons by which it can mold the thoughts, desires, and wills of its citizens; nor is it really concerned to exert such influence. It does not seek to make men truly good or virtuous... the state is a non-natural, remedial institution... it is a consequence of the Fall. It is both a punishment for sin and a remedy for man's sinful condition; without it anarchy would reign, and self-centered, avaricious, power-hungry, lustful men would destroy one another in a fierce struggle for self-aggrandizement" (Deane, 1963, p. 117). Alasdair MacIntyre's claim that there is something "radically new" in Augustine's theology is essentially right: "Augustine of course was interpreting the scriptures in his accounts of human sinfulness and redemption, but his elucidation of them is distinctive in that he is able to give an account of Paul's doctrine in *Romans*, for example, using a vocabulary of a kind which was not available to Paul himself" (1988, p. 157). Porter's reading Augustine's political theology into Paul is one way of understanding Paul's texts, but it is very different from that of the consensus view.

⁵ Finnis is not alone among contemporaries who challenge the transmission principle. John Hittinger objects that the "Thomistic warrant for [the transmission principle] is quite slim" (2002, p.

LYCEUM

philosophers who advocated the ‘Pauline principle’ were guilty of legalism. Lawyers naturally consider the original right of possession in any transaction where present rights of possession are in doubt. The transmission theorist similarly looks for some original right that is later transferred to another, but for Finnis no such original right can be discovered because none exists. That doesn’t matter much to political authority because, Finnis argues, its grounding lies elsewhere: political authority is grounded in and justified by the need for a solution to coordination problems that are unavoidable in any social group, including political communities.

What has happened? In order to answer this question in a provisional way I return to Maritain’s exposition of the ‘Pauline principle’ in *Scholasticism and Politics* and other works. I use Maritain’s works as a starting point to flesh out an account of the two principles of the consensus view, but I devote priority in the

40). As he points out, the principle depends crucially on remarks made by Aquinas in the *Summa Theologiae*’s “Treatise on Law” (I-II 90, 3). He concedes that Maritain and Simon, in relation to whose work Hittinger takes up the principle, “are following the lead of the great Thomistic commentators Cajetan, Bellarmine, and Suarez all of whom use this text.” For Hittinger, however, neither the text in Aquinas, which he describes as an “ambiguous passage” (though he never explains what is ambiguous about it, nor does he acknowledge that this is not the only passage used to establish the transmission principle). Hittinger’s discussion is focused on the defense of liberal democracy in the works of Simon and Maritain, and he can be faulted for conflating a defense of democratic philosophy with a defense of democratic practice. Maritain and Simon both defend the former, but neither is committed to the universal and exclusive validity of the latter. It is worth noting in this regard that a democratic philosophy can and should inform other forms or modes of practice, including monarchy and aristocracy. This failure, I believe, accounts for the further mistakenness in Hittinger in identifying “consent through election” as a primary means of transmission. Voting is not a sufficient indicator of voluntary acceptance of a state (that is rarely if ever what elections are about), and, even if it were, conceptualizing transmission in terms of consenting to being governed runs the risk of making the theory indistinguishable from standard consent theories of political authority, but in such accounts authority isn’t derived from God but rather from the voluntary acts of human beings who choose to live together. Besides, if it were the case that consent of the governed is the only acceptable means of transmission, then the transmission theory would be exclusively tied to democratic practice, and it would be very difficult to link the same back to Aquinas. I do think, however, that on occasion Maritain and Simon are not sufficiently clear about their distinction between democratic philosophy and democratic practice.

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

first part of the paper to evidence for the consensus view in the works of Aquinas since all of its defenders were committed to its fidelity to Aquinas and the natural law tradition of political philosophy. I also examine in greater detail Finnis' critique of the theory, which requires some exposition of his own version of the natural law's defense of political authority. I take this up in the second part of the paper. The goal of the paper is to determine whether or not the two principles of the consensus view ought to be recovered as a viable component in a Thomist justification of political authority, as both the figures who gave the theory its first exposition in the early modern period—Bellarmine and Suarez—and those in the first half of the twentieth century conceived it to be.

1. The Consensus View

In this part of the paper I lay out some of the evidence for the two principles of the consensus view. I start with and return to Maritain, but my focus will be on the available textual evidence in Aquinas. My aim is threefold: first, I want to describe the two principles in somewhat greater detail; second, since the consensus view has fallen into neglect, I want to point to explicit evidence for it in Aquinas as well as later thinkers in the natural law tradition who endorsed the consensus view; and, third, since the chief reason for seeking to recover the consensus view would be its truthfulness and adequacy as an account of political authority from the natural law tradition, I need to provide a sufficient account of the view so as to enable me to respond to its most vocal critic among contemporary natural law thinkers.

Supporters of the consensus view take as their starting points not only a natural law conception of political morality, but also an eternal law framework within which the natural law finds its ground and justification. Maritain's reference to the Pauline principle in *Scholasticism and Politics* is only one of several similar references. In *Man and the State*, for example, Maritain (1951) connects the "majesty" of political authority with the fact that the authority

LYCEUM

represents the people, which representation is justified in terms of the Pauline principle: “since it represents the people, the civil power holds its authority, through the people, from the Primary Cause of Nature and of human society. St. Paul teaches that ‘there is no authority that is not from God’” (p. 131).⁶ In *Ransoming the Time*, Maritain (1972a) writes: “Saint Paul requires those under authority to obey those above. For him it is a duty of man’s conscience to have respect for legitimate authority, because all legitimate authority has its primary source in God, Author of nature” (p. 201). Several key features of the Pauline principle are enunciated by Maritain: (1) political obligation is binding in conscience—it is a moral duty of subjects; (2) God is the source of such an obligation inasmuch as God is the source of authority itself; and (3) it is because God is the creator (“Author” and “Primary Cause”) of human nature that such a nature requires political authority and that God is the source thereof—that is, the act of creation gives to human beings a particular set of inclinations that can only be realized fully within the types of community that require political authority, and the same act of creation gives to God the power to bind the consciences of rational creatures.

At the heart of Aquinas’ natural law theory in the *Summa Theologiae* lay the same set of claims. Maritain attributes both the Pauline and transmission principles to Aquinas: “In the political order . . . , the organs of government are then regarded . . . as having the source of their authority in God. . . . Once the organs are designated, authority resides in them . . . whose personification and vicar they are: *vices gerens multitudinis*, as St. Thomas puts it” (1996, p. 278). It is possible, however, to conceive of natural law theory as independent of the

⁶ This should be kept in mind in reading later passages of the work, like the following: “the foundation of the Church as a society was accomplished from above downwards, but political society originates from below upwards. In other words, authority in the Church comes down from above, but authority in political society rises from below; and whereas the Pope in the Church is the Vicar of Christ, the rulers in political society are the vicars of the people” (184-5). The authority transferred from the people to their political rulers, however, comes ultimately from God: Maritain’s point is merely that this same authority, an authority that is vested in the people by God, is transferred by the people.

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

findings of natural or philosophical theology. Finnis (1980) appears to have taken just such a stance in *Natural Law and Natural Rights*: “Part II of this book offers a rather elaborate sketch of a theory of natural law without needing to advert to the question of God’s existence or nature or will” (p. 49).⁷ Similarly, in his *Aquinas*, despite the potential “suspicion that Aquinas assimilates the government of free people too closely to the ordering of physical and biological nature,” Finnis asserts that “when developing his ideas more philosophically and without the pressure to make use of every traditional theological form of speech, Aquinas strongly insists that law is something addressed by one mind and will to others—by one freely choosing person *to other freely choosing persons*.” As such, this “firmly relegates” the eternal law to a “non-focal” meaning of law in Aquinas in contradistinction to the “central case of law” which is “an appeal to the mind, the choice, the moral strength {*virtus*}, and love of those subject to the law” (p. 49).⁸

⁷ It would be unfair not to mention that Part III of the book takes up the relations of nature, reason, and God, but it isn’t clear from what is said there that Finnis retracts his earlier claim. In his contribution to the “Founders of Modern Political and Social Thought” series, *Aquinas/Moral, Political, and Legal Theory*, Finnis again reserves his reflections on the place of God in the Thomistic system until the final part of the book. There Finnis seems to set up a dichotomy based on the audiences of Aquinas’ works. On the one hand, in writing for theological students, Aquinas is said to be “willing to speak of an ‘eternal law’” (2004, p. 307). On the other hand, “when developing his ideas more philosophically and without the pressure to make use of every traditional theological form of speech, Aquinas strongly insists that law is something addressed by one mind and will to others—by one freely choosing person *to other freely choosing persons*. This does not contradict the idea of an eternal law governing even subrational creatures, but it firmly relegates that idea to an extended, non-focal sense of ‘law’. The central case of law is an appeal to the mind, the choice, the moral strength {*virtus*}, and the love of those subject to the law.”

⁸ In his *Aquinas/Moral, Political, and Legal Theory*, Finnis again reserves his reflections on the place of God in the Thomistic system until the final part of the book. There Finnis seems to set up a dichotomy based on the audiences of Aquinas’ works. On the one hand, in writing for theological students, Aquinas is said to be “willing to speak of an ‘eternal law’” (2004, p. 307). On the other hand, as suggested by the passage quoted, when Aquinas writes for what appears a more philosophical audience—at least this is how I am interpreting the ambiguous claim about the

LYCEUM

Finnis' position is difficult to square with central claims of Aquinas' natural law theory, however. To begin with, Aquinas argues that in any sequence of causes, the effect depends more on the first cause than on subsequent causes. This general claim is made prior to the particular assertion that reason is a rule of human action only insofar as it is informed by and conformed to the eternal law: "It is therefore evident that the goodness of the human will depends on the eternal law much more than on human reason" (*ST I-II 19, 4*). Two questions later Aquinas will refer to this argument as the authority in the *sed contra* of the first article in order to establish that "the goodness of a human action depends principally on the Eternal law" (21, 1*sc*).⁹ The priority of eternal reason to human reason is a commonplace of Aquinas' moral theory: "there are two rules of the human will: one is proximate and homogeneous, viz. the human reason; the other is the first rule, viz. the eternal law, which is God's reason" (*I-II 71, 6*). It is unclear to me, therefore, how the focal meaning of law can be human law.

Human laws have the power to bind the consciences of those subject to them from their connection to the eternal law (*ST I-II 96, 4*). Aquinas states in terms of an exceptionless disjunction that all laws are either just or unjust. All just laws are said to derive from the eternal law.¹⁰ As "a dictate of practical

pressure Aquinas felt to use all forms of theological discourse in some cases but not others—he shifts the focus away from a theological ground for natural law.

⁹ Aquinas goes on to argue in the body of the article that, given the dependence of the ordering of an agent to an end on a rule or measure, in human action "the proximate rule is the human reason, while the supreme rule is the Eternal Law." The rightness and wrongness of human actions are ultimately determined by conformity with the eternal law.

¹⁰ In *I-II 96, 4* Aquinas cites as his scriptural warrant for this claim another key commonplace, though this one comes from the Old Testament *Book of Proverbs*: "By me kings reign" (8:15). In the *sed contra* to 93, 3 Aquinas again refers to this textual warrant to support the claim that *all* laws derive from the eternal law. The body of the article makes the dependency explicit, again drawing on the efficacy within sequential causation: "the law denotes a kind of plan directing acts towards an end. Now wherever there are movers ordained to one another, the power of the second mover must needs be derived from the power of the first mover... Wherefore we observe the same in all those who govern, so that the plan of government is derived by secondary governors from the governor in chief; thus the plan of what is to be done in a state flows from the king's command to his inferior administrators... Since then the eternal law is the plan of government in the Chief

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

reason emanating from the ruler who governs a perfect community,” the *ideas* in the “Divine Reason”—which governs “the whole community of the universe”—provide the source and model for natural and human law (*ST I-II 91, 1*). Indeed, all law derives from and receives its power to bind from the Eternal Law. Eternal law governs the whole universe; the natural law governs human conduct in general, but it requires greater specificity to apply to human interaction in social groups, and that specificity is given in human “positive” law (*ST I-II 99, 3ad2*). Against Finnis, therefore, I follow Alasdair MacIntyre (1988), who contends that “perfected obedience to the natural law requires the virtue of justice in full measure,” which must include religious practice—since religion is a moral virtue affiliated with the virtue of justice. Affiliation does not mean subordination: for Aquinas religion is the most important of the moral virtues—that is, it ranks higher in his estimation of the moral virtues than each of the cardinal virtues because its act is nearer to the end of all of the moral virtues (*ST II-II 81, 1*).¹¹ How could one possess the virtue of religion, a requirement of the natural law, without believing in God? MacIntyre concludes that one cannot: “It is then important to my interpretation of Aquinas’ positions that I understand his positions on practical knowledge and practical reasoning, let alone those on justice, as always presupposing the type of rational knowledge of God exemplified in the conclusions of the *Prima Pars*” (p. 188).¹²

Governor, all the plans of government in the inferior governors must be derived from the eternal law. But these plans of inferior governors are all other laws besides the eternal law.”

¹¹ The *sed contra* of the article provides insight into the interconnectedness of law and virtue in Aquinas’ thought as well as the source of both in God: “The precepts pertaining to religion are given precedence as being of greatest importance. Now the order of precepts is proportionate to the order of virtues, since the precepts of the Law prescribe acts of virtue.”

¹² *WJWR*, p. 188. MacIntyre is equally explicit about his commitments to the philosophical theology of Aquinas as central the latter’s moral and political philosophy in his later works as well: “Modern Catholic protagonists of theories of natural law have sometimes claimed that we can fully understand and obey the natural law without any knowledge of God. But according to Aquinas all the moral precepts of the Old Law, the Mosaic Law summed up in the Ten Commandments, belong to the natural law, including those which command us as to how we are to regard God and comport

LYCEUM

The same privileging of the eternal law provides an occasion for Maritain to add what might be called a *legitimacy condition* to political obligation: the obligation to obey is circumscribed by principles of justice that must be independent of political authority. The principles of justice must be independent of political authority to avoid circularity: otherwise all law would be just merely because a respective authority laid claim to their being so. For Maritain (1972b), as for Aquinas, the relevant principles of justice are derived from the eternal law: “[Aquinas] held that laws are only truly law because they are just, and that they are binding in conscience only if they are just. A real justice, not a feigned one, is the foundation for authority in law. . . . If we shatter this basic order, which links things human to the divine stabilities of the universe, even the strongest empirical defenses of the social order will remain vain” (p. 37). The dependence of all law on the eternal law provides the objective, exterior measure required of human laws—that is, laws that require the violation of an eternal or natural law are unreasonable and as such lose their power to bind in conscience (I-II 96, 4).

Maritain’s remarks immediately preceding the naming of the Pauline principle in *Scholasticism and Politics* connect the principle to Aquinas’ political naturalism. The reason for making the connection has to do with the order of discovery. It is with the necessity of political association for the flourishing of human beings that we begin the study of the origin and source of political authority because it is with this that we are most familiar. Maritain argues:

ourselves in relation to Him. A knowledge of God is, on Aquinas’s view, available to us from the outset of our moral enquiry and plays a crucial part in progress in that enquiry. And it would be very surprising if this were not so: the unifying framework within which our understanding of ourselves, of each other, and of our shared environment progresses is on in which that understanding, by tracing the sequences of final, formal, efficient, and material causality, always refers us back to a unified first cause from which flows all that is good and all that is true in what we encounter. So in articulating the natural law itself we understand the peculiar character of our own directedness, and in understanding the natural law better we move initially from what is evident to any plain person’s unclouded moral apprehension to what is evident only or at least much more clearly to the *sapientes*, those whom Aquinas saw as masters of the master-craft(I-II 100, 1), and to what supernatural revelation discloses” (1990, p. 141).

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

If in the cosmos, a nature, such as human nature, can only be preserved and developed in a state of culture, and if the state of culture necessarily entails a certain condition—the relation of authority among men—this relation is demanded by natural law. (p. 103)

The argument consists of two conditionals: *if p then q, if q then r*. The two conditionals appeal directly to the naturalist justification of political obligation that Aquinas adopted from Aristotle. The first conditional—the dependence conditional—asserts that human development (specifically, human flourishing) is dependent on participation and membership in a political community. Maritain's "state of culture" may be misleading if it is thought that such culture is possible outside or independently of political organization. If it is understood as synonymous with political community (or at least dependent thereon) inasmuch as the type of perfection that rational beings are capable of requires the existence of a culture that can only be developed out of and nourished within a political community, the ambiguity can be avoided.¹³ The second conditional is not expressly stated by Aristotle. One might argue that the first implies the second: if political community is a necessary condition of human flourishing, then, assuming that human flourishing itself is the good for human beings, whatever is necessary for political community is also necessary for human flourishing. Such

¹³ There remain at least two potential objections to the dependence conditional: (1) it can be argued that human beings do not have a nature—the anti-essentialist objection; or (2) that even if such a nature exists and can be identified as such, it can be argued that there is no dependence of this nature on political community. This is not the place to address these objections, but the first will apply to any teleological eudaemonist ethic and the second to every naturalist defence of political authority. Each objection, therefore, threatens the foundations of the tradition of political naturalism because both entail a rejection of the starting points of that tradition. It is not surprising that early modern political philosophers sought to replace at least one of these starting points: inasmuch as the state of nature device in Locke's contract theory, for example, presupposes the possibility of pre-political human culture it sets itself in direct opposition to the dependence condition.

LYCEUM

at least is the logical structure used by Aquinas explicitly in defending the second precept of charity—that is, love of neighbour—in the *Summa Contra Gentiles*: “Besides, since ‘man is naturally a social animal’, he needs to be helped by other men in order to attain his own end. This is most fittingly accomplished by mutual love which obtains among men. Therefore, by the law of God which directs men to their ultimate end, mutual love is prescribed for us” (SCG 3, II, c. 117.4).¹⁴ Just as political authority is said to be required by the natural law, so mutual love is defended as required “by the law of God” because it is necessary to successful political cooperation.

Maritain’s double-conditional also echoes one of the clearest statements of Aquinas’ naturalism in the *Summa Theologiae*. Authority would have existed in the state of innocence, Aquinas asserts in the *prima pars* “because man is naturally a social being, and so in the state of innocence he would have led a social life. Now a social life cannot exist among a number of people unless under the presidency of one to look after the common good; for many, as such, seek many things, whereas one attends only to one” (96, 4). As we will see, Finnis himself makes much of this claim by Aquinas. For Maritain (1940), what matters is that the grounding in the natural law of authority points inevitably to the source of that law and its grounding in human nature. Maritain points to a “twofold meaning” in this grounding (p. 104). First, “hierarchical differentiations”—that is, the relationship of ruler and ruled, of authority and obedience—are required by the very nature of social life. Second, given the natural equality of all humans as members of the same species, the general right to govern others requires a special justification; indeed, it is possible “only if nature itself is considered, not as a simple collection of phenomena, but as the work and the created participation of a supreme ordinating Law, ‘justified in

¹⁴ In the *Summa Theologiae*, Aquinas will use a similar strategy for the inclusion of justice in the moral order of human action. If human beings were not political by nature, then human actions would need only conform to the dictates of reason and of God; but, because human beings must live together in communities to flourish, their actions must also conform to the dictates of shared living, including the virtue of justice and, we can add from the passage in the *SCG* the demands of love (ST I-II 72, 4).

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

itself” because identical with the absolute Good” (p. 104-5). The justification of the right to govern, therefore, presupposes the existence of the eternal law and eternal reason—that is, it presupposes that all authority comes from God whose reason is the source of the eternal law.

1.1 The Pauline principle

The Pauline principle holds that political authority or the right to direct and command is derived from and has its origin in the Creator, as “supreme ordinating Law.” Aquinas not only held that human government was derived from the divine, but also that it should imitate its source (*ST* II-II 10, 11). Not surprisingly, the most complete statement of the Pauline principle in Aquinas comes in his commentary on *Romans* 13:1-7, though it is surprising that this text is not regularly cited by most of the mid-twentieth century defenders of the consensus view. The fact is that it is one of the two principal New Testament sources that Aquinas prefers to cite when discussing political authority and the duty to obey.¹⁵ The first two verses from Paul’s letter express the core of the principle: “Let every soul be subject to the higher powers. For there is no power except from God, and those that exist have been instituted by God. Therefore he who resists the authorities resists what God has appointed, and those who resist will incur condemnation.” It is in his commentary on these verses that Aquinas articulates clearly his own commitment to the Pauline principle. According to Aquinas, Paul is addressing early Christians who believed “that they should not be subject to earthly powers on account of the freedom they received from Christ” (Aquinas, 2012, §1017). On the contrary, Paul admonishes the community to believe that, during this life, subjection to authority is owed by all,

¹⁵ The second text is *Hebrews* 13:17. In comparison, another commonplace text, *1 Peter* 2:18, Aquinas only refers to once in the *Summa Theologiae*, and only 3 other times in his entire corpus in each case in a scriptural commentary.

LYCEUM

in Aquinas' words, "according to the order of justice."¹⁶ Aquinas goes so far as to argue that Paul avails himself of the indefinite expression, *higher powers*, in order to claim that obedience is owed to "the sublimity of their office" rather than the person in office, which means that subjection extends even to those that are wicked rulers (§1018). With this "admonition" to the early Christians in mind, Aquinas proceeds to explain how the remainder of these early verses establish that subjection to political authorities is both virtuous and necessary (§1020).

To establish that subjection is virtuous, Aquinas advances two arguments. First, he provides Paul's simple assertion that all power is from God with syllogistic support via the following premises: (i) when something is predicated of God and creatures it must come to creatures from God, and (ii) power is predicated of God and human beings. From these premises Aquinas concludes that "all human power is from God" (§1021). Then he considers an objection to the assertion that power originates from God by referring to *Hosea*: "They made kings, but not through me." In response, he distinguishes three facets of power: its origin, acquisition, and use. Having already established that its origin is God, Aquinas simply cites another scriptural authority to this effect. Texts like the one from *Hosea* must be understood to refer not to the origin of authority, about which Aquinas no longer entertains any doubts. Instead, the referent of such texts must be either the acquisition of power or its use. In both of these cases it is possible for one to claim that the power does not come from God. On the one hand, when the person who acquires power does so illegitimately, through "perverse desire" for example, we can claim that the power is not from God. Similarly, it is possible to disclaim that power is from God when the person using power violates "the precepts of divine justice" (§1022). Only the just acquisition and use of power can be said to come from God in a certain way of speaking, but speaking strictly about its origin all human power is caused by God.

¹⁶ "*secundum iustitiae ordinem*"

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

Aquinas takes that conclusion as the first premise of the second argument to prove that subjection is virtuous. To it he adds another to the effect that what is from God is well-ordered. Together these are taken to prove that “the order whereby the lower are subjected to the higher powers is from God” (§1025). Resistance to such authority constitutes resistance to the “divine order” of things, which is “contrary to the good of virtue.” Subjection to authority is consistent with the divine order and required by the origin of political authority. This also means, however, that all such power is subordinate to and at the service of divine authority (§1028 & 1034).

It would be something of an understatement to say that there is ample evidence elsewhere in Aquinas' *corpus* for the Pauline principle. It is evident in early and later works. In the *Commentary on the Sentences*, Book II, Distinction 44, Question 2, Article 2, Aquinas not only cites *Romans* 13:2 in the *sed contra*, but also draws more fully on the text in articulating his defense of the political obligations of subjects to authority:

I respond by saying that, as has been said, obedience looks to the obligation of observing in the command which is served. However, this duty is caused by the order of precedence which possesses constraining force, not only temporally but also spiritually as a matter of conscience, as the Apostle says. Insofar as the order of precedence derives from God, as the Apostle indicates in the same place. And thus, insofar as what comes from God, the Christian is required to obey such people, but not insofar as the precedence is not from God. (Malloy, 1985, pp. 161-64)¹⁷

¹⁷ On three occasions in the text I have substituted “precedence” for Malloy’s preferred “sovereignty” as a translation of *prelatio*. One reason for doing so is to indicate that though it is used throughout the *Sentence* commentary, *prelatio* is less common in relation to political authority

LYCEUM

What follows closely anticipates the arguments laid out in the commentary on *Romans*. He indicates, for example, that we can speak of power failing to come from God whenever its acquisition or use fails to meet certain conditions. On improper acquisition Aquinas includes two categories: the worth of the person holding office and the manner through which they came to power (i.e. violently). Only the latter releases one from an obligation to obey. The former does not because worth does not prevent one from properly acquiring office,¹⁸ and so the fact that even unworthy individuals exercise authority does not excuse subjects from their obligation to authority because “its form is always from God, and this creates the duty of obedience, the subjects are thus required to obey such superiors, though they may be unworthy.” Both of the conditions related to misuse excuse: (1) rulers who command what is “contrary to that for which the sovereignty is established, as if it commands an act of sin contrary to virtue” and (2) rulers who exceed the limits of their proper jurisdiction may be disobeyed.

To return to the question on human law in the *Summa*, for a moment, Aquinas undertakes to respond to an objection that human law cannot bind in conscience because “an inferior power has no jurisdiction in a court of higher power” (*ST* I-II 96, 4 obj. 1). But what could be higher than the divine that governs the consciences of human beings? In his response Aquinas cites *Romans* 13:1-2. Indeed the entire argument is little more than a restatement of claims that he had made in the *Romans* commentary: since “all human power is from God” those who resist authority that is exercised within its proper jurisdiction are culpable in their conscience of violating the order of God (*ST* I-II 96, 4*ad*1). An objection to the right to appeal cites the *Romans* text as support for its position:

later works. Another is that it avoids some of the connotations that have accrued to sovereignty in the early modern period.

¹⁸ The point is not that Aquinas believes unworthy individuals are indistinguishable from worthy ones. The point is narrower: even unworthy individuals can come to power through the appropriate procedures, even when these procedures are normally effective in rooting out the incompetent or immoral. Aquinas, of course, no doubt has monarchical succession in mind, and so it is even easier to see that unworthy individuals could come to office through accepted means such as primogeniture.

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

given that we are enjoined to submit to authority, appealing judgments constitute a form of sin by undermining the authority of the judge. Against this position, Aquinas defends the right to appeal wherever “the lower authority departs from the authority of the higher” (*ST II-II 69, 3ad1*). Aquinas has in mind cases where judicial authority is exercised in bad faith, resulting in the unjust oppression of those who find themselves before such a judge. But at *ST II-II 69, 4*, Aquinas uses the Pauline principle in the *sed contra* to argue against resistance to capital punishment where the sentence is *justly* arrived at. Self-preservation does not trump obedience to political authority. And, in the discussion of disobedience as a vice, the Pauline principle is invoked against “disobedience to the commands of a superior” insofar as resisting lawful authority is equivalent to resisting the order of God (*ST II-II 105, 1*).

One finds an even more informative use of the principle in the fifth of six articles devoted to the virtue of obedience. In this article, Aquinas argues that political authority is not unlimited: the obligation to obey does not extend to commands made by a superior that violate Divine commands. My interest in this article, however, is not with its key point but with the manner in which Aquinas clarifies the obligatory force of the Pauline principle. To do so, he likens the “necessity of justice”—through which one is bound simply to Divine commands and in a limited way to the commands of a political superior—to “natural necessity” (*ST II-II 104, 5*). In the normal course of events, natural things are moved by their movers. The normal course can be thwarted, however, in two ways: (i) it can be prevented by a stronger mover; or (ii) there can be that in the thing moved that is not wholly determinable by the mover. These limits on natural necessity have analogues in the necessity of justice: (i) Aquinas rehearses an ascending series of superiors with God at its apex such that each subordinate command is circumscribed by that which is higher up until one reaches the apex; and (ii) even though human subjects are obligated to obey the commands of their political superiors, these commands do not extend to the internal acts of a human being, only the commands of God do so. God’s authority is unique in this way

LYCEUM

because it is not only mover but cause: “it is possible for something to move a natural thing, without being the cause of the thing moved, yet that alone, which is in some way the cause of a thing’s nature can cause a natural movement in that thing” (*ST I-II 9, 6*). Aquinas gives as an example of this premise the tossing of a stone into the air. By such an action the stone is moved, but the movement is not natural to the stone. Only the *cause* of the stone’s nature can move it in a natural way. The same holds for the wills of human beings, the cause of which is God.¹⁹ If political authority were not derived from divine authority, its ability to move human beings by commanding them would amount to violence in the sense that the movement would be opposed to the natural movement of the will. For that movement to be natural it must be derived from its cause. That it is so derived is the claim of the Pauline principle. Political authority stems from and seeks to return people to God, but its scope is necessarily narrower than and its force necessarily inferior to that of divine authority from which it is derived: “inferiors are not subject to their superiors in all things, but only in certain things and in a particular way, in respect of which the superior stands between God and his subjects” (*ST II-II 104, 5ad1*).

1.2 The transmission principle

Having set out some of the evidence for the Pauline principle, I turn to the second principle of the consensus view, the transmission principle. Granted that Aquinas traces the source of all authority to God, does he also share the view that the transmission of this power is mediated through the people to those in positions of authority? I wish to provide some context for the principle before turning to the textual evidence.

¹⁹ Aquinas provides two reasons for believing that “the cause of the will can be none other than God”: (1) as a power of the rational soul the will is caused by God inasmuch as the rational soul is and (2) because “a particular cause does not give a universal inclination,” and because the will is inclined to the universal good it must be caused by that which is universal in essence rather than by participation but only God is universal in essence.

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

In the late nineteenth century two subordinate principles of political authority competed for supremacy in Catholic political philosophy. For both the source of political authority and the correlative obligations of citizens to obey the laws was God—that is, both take as the appropriate starting point the Pauline principle.²⁰ The differences between the two principles hinge on the manner in which the power of binding the consciences of human beings passed from the Creator to one or several members of the human community. One principle argued that the transfer of authority was unmediated, which meant that specific political authorities could trace their power directly to God. God transferred the authority directly to the person or persons in power, as designated by the people. On this principle, the people only act as an “instrumental cause” of authority (McCoy, 1989, p. 50). This principle arose in the nineteenth century in response to the fragmentation of political communities that was blamed on the contract models of political society that had developed in the seventeenth and eighteenth centuries. It was feared that the contractual model would lead to further disintegration, and the French revolution was no doubt an important catalyst for this way of thinking. Known as the designation principle, it was much newer than the transmission principle which is usually traced back to Aquinas. Proponents of the older principle hold that political authority is received by the person or persons in authority in a mediated manner. The idea is that authority is given by God to the whole people of a political community, who in turn transfer or transmit the authority in question to a political leader or group of leaders. It is in and through this effort of transmission that the people, not individual citizens,

²⁰ Note that both theories accepted that the differentiation of regime types—that is, into monarchies, aristocracies, democracies, and so on—was a separate and subordinate issue. Each of the principles sought to justify the foundation of political authority; neither was necessarily concerned with the mode of its functioning. This point is important because even if the people were to exercise some form of direct democracy, such that no distinct governing personnel were identifiable, an account of the authority of the majority to bind the consciences of all members of the community to its dictates would still be required.

LYCEUM

can be understood to be the direct cause of political authority. Of course they remain secondary to the primary cause of political authority.

Consider in the light of this the following texts from Aquinas. Both texts come from the “Treatise on Law” in the *Summa*. Let me begin with the second. Question Ninety-seven deals with changing laws, about which Aquinas is sensibly cautious given the potential threat to regular obedience. In Article Three, Aquinas addresses the issue of whether or not custom can acquire the force of law. He will argue that it can, and in doing so faces the following objection. Making law belongs to the person or persons charged with the governance of the community; such individuals act in making laws in a public capacity; but custom is made through the repeated acts of individuals in their private capacity as subjects or citizens. In response, Aquinas distinguishes between two conditions of people: they may be free citizens capable of making laws or they may not be so situated as to make their own laws. It is not entirely clear what condition of people Aquinas has in mind in the first instance, though given his preference, articulated elsewhere,²¹ for a mixed regime, I assume he has in mind some shared power arrangement whereby the citizens of a given jurisdiction are empowered to make laws even though they also have a prince who both makes and enforces laws. Despite the difficulties with understanding the context, it is relatively clear that in such a situation the people would be acting in a public capacity inasmuch as their customs would only take on the force of law where the “whole people” participated in the observance, not just any individual. For the second group, although they are not capable of acting in such a public capacity, their customs can also take on the force of law only if they are “tolerated by those to whom it belongs to make laws for that people: because by the very fact that they tolerate it they seem to approve of that which is introduced by custom.”

Recall that the transmission principle holds that the people are real, though secondary, causes of political authority. Advocates of the principle see in Aquinas’ defence of customary law an articulation of the principle. In the first

²¹ See ST I-II 105, 1. See also James M. Blythe (1986) and Mark R. MacGuigan (1957).

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

instance, on my reading, we have a situation that is relatively close to a direct democracy, and in such a situation the source of authority still calls for some explanation. It is explained, I think, by the essential possession by the whole people of the right of government. This is equally true in the second case where authority is exercised by a distinct governing personnel.²²

The second passage comes earlier in the “Treatise on Law” in Article Three of the First Question on law. The First question treats law in general, and it concludes by picking out four essential features of law: that it be of reason, for the common good, made by someone in authority, and made available to the people. In the Third Article Aquinas argues that law must be made by someone in authority. The argument begins with the claim that the purpose of law is to direct others to the common good, and Aquinas argues that this directive or ordering capacity “belongs either to the whole people, or to someone who is the [vicar] of the whole people” (*ST* I-II 90, 3). Given the generality of the law and the necessary freedom of those members of a political community who can be considered citizens, laws must only be given to “the people” (I-II 98, *6ad2*). What constitutes “the people”?

These passages from the *Summa* do not adequately characterize “the people,” and in Aquinas’ political theology “the people” are never afforded an exhaustive treatment. Yet there are helpful signs in the relevant texts. For example, the very generality of law, which is a condition of law as opposed to the particular judgments of individual rational agents, requires that it is applicable only to a social group of sufficient magnitude and only as a whole: “A law should not be given save to the people, since it is a general precept” (I-II 98,

²² The locution may appear odd, but it preserves a key distinction. Although all authority comes from God and is justified in light of the eternal and natural laws, the choice of the form of government is left open to human choice—that is, the consensus view, including Aquinas, leaves open the possibility that different regime-types will be better suited for different peoples and even for the same peoples at different times in their history. There is sometimes a tendency to conflate the transmission principle with direct democracy, but that isn’t required by the principle and clearly is not a preferred option for Aquinas.

LYCEUM

6ad2). On two different occasions, moreover, Aquinas avails himself of Augustine's re-conceptualization of the Ciceronian definition of the *people* in the *City of God*: "the people . . . is an assemblage associated by a common acknowledgment of right and by a community of interests" (Augustine, 2000, 19.21). Aquinas' reflections on the Augustinian definition come from his *Commentary on Hebrews* and the *Summa Theologiae*, respectively: (1) "When therefore they consent to the right of divine law, such that they are useful to each other and tend unto God, then there is a people of God" (§406) and (2) "it is of the essence of a people that the mutual relations of the citizens be ordered by just laws" (*ST* I-II 105, 2).²³ In both instances, what is essential to the formation of "a people" for Aquinas is that they conform themselves to the laws, whether the laws of a religious or a political community. Surely conformity to the laws is only a necessary condition for defining a people. As a result of their conformity, they will have several features in common despite the intrinsic diversity of the members of the community, and their common features are just those attributes that can be shared or communicated with each other in friendship, which can only exist among rational beings (*ST* I 20, 2ad3). Indeed, human law has as its "principal intention" the production of friendship among citizens (I-II 99, 2).²⁴ Friendship is necessary for flourishing in this life, even for the flourishing of contemplatives (I-II 4, 8). The members of a political community will be encouraged by the laws to develop the types of virtuous characters that enable them to befriend each other, "since every law aims at establishing friendship,"

²³ In the second passage I have substituted "people" for the translation's "nation" as closer to the meaning of Aquinas' *populi*.

²⁴ The Old Law had the same intention: "all the precepts of the Law, chiefly those concerning our neighbor, seem to aim at the end that men should love one another," and it is to further this end that the Law permitted a guest to eat the fruit of a neighbour's orchard—so long as nothing was taken away (the eating had to be done onsite, as it were)—because such a policy "strengthens friendship and accustoms men to give things to one another" (I-II 105, 2ad1). Again, given that the Law aimed to have members of a political community ready to assist one another—"a very great incentive to friendship"—it also provided incentives to encourage lending and discourage burdensome debt recovery practices.

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

either between members of the community or between the same and God (I-II 99, 1ad2). Theirs will be a common set of goods whose realization depends on their on-going participation in practices that instantiate the relevant goods. In more perfected communities, citizens will have made the transition from merely dealing with each other in terms of legal obligations to moral obligations, including those of friendship.²⁵ It is from within such communities that the transmission occurs.

Let me return to the transmission principle itself. In *Man and the State*, Maritain argues that there is a need to “sharpen the philosophical concepts” that we use to understand the origin of authority, and he has at least one of these two passages from Aquinas in mind (p. 133). What Maritain seeks to clarify is the concept of *vicariousness*. He distinguishes between two manners in which one may be said to give something to another. The following text from Aquinas’ *An Apology for Religious Orders* points toward the distinction that Maritain wishes to make:

The proposition, that what a man gives away he does not still possess, does not hold good in things spiritual. These are communicated, not like physical things, by the transference of some dominion over them, but rather by an emanation of an effect from its cause. When one man communicates knowledge to another he does not, on this account, deprive himself of this knowledge; for it remains in his power. In the same way, he that confers some power

²⁵ “The Philosopher does not deny that friendship is a virtue... For we might say that it is a moral virtue about works done in respect of another person, but under a different aspect from justice. For justice is about works done in respect of another person, under the aspect of the legal due {*debiti legalis*}, whereas friendship considers the aspect of a friendly and moral duty {*debiti amicabilem et moralis*}” (II-II 23, 3ad1). For further examples of this distinction between the legally and morally obligatory in relation to political life, see Aquinas’ remarks on gratitude (II-II 106, 5), revenge (II-II 108, 2ad1), and liberality (II-II 118, 3ad2).

LYCEUM

upon another, does not, by so doing, deprive himself of that power. (p. 135)

On the one hand, we can speak of giving to another in a material manner, such that the giver relinquishes possession of a material thing at the same time that the receiver takes possession. Take as an example the act of giving a friend one hundred dollars. In this sense there is a real separation of the thing from the giver. At the end of the transaction, the friend has, as the giver does not, the money. On the other hand, we can speak of giving to another in a *vicarious* manner. In such giving the giver retains possession of the “moral or spiritual quality” that is given to the receiver. To explain this second type of giving Maritain appeals to the notion of participation: the receiver participates in the “moral or spiritual quality” that is given, but the gift remains with the giver as belonging to his or her essence. The example he gives makes the point much more effectively: the student participates in the knowledge that belongs to the teacher essentially by right of tuition, and yet in providing this gift of knowledge to the student the teacher does not lose the same.

Applied to the principle of transmission, Maritain (1951) holds that the concept of vicariousness provides a fuller account of the transfer of authority from the people as cause in two senses: “The first relates to the fact that in investing rulers with authority the people lose in no way possession of their basic right to self-government. The second relates to the fact that the representatives of the people are not mere instruments, but rulers invested with real authority, or right to command” (p. 134). The distinction aims to avoid two errors. It would be an error to treat the transmission of authority as a material transmission, as in the case of giving a friend money, because this would treat the transmission as final. If nothing else, such a conception runs into problems with future generations who were not party to the original transmission. How is it that the present generation could consign the rights of future generations in such a manner? It would equally be an error to treat the transmission in a way that the authority transferred was only exercised in an instrumental fashion at the

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

discretion of the people. This would really amount to no exercise of authority whatsoever: the person or persons in authority would not be directing others but being directed by them. Their authority would be that of a proxy.²⁶ To return to the example of the student, should the student him or herself become a teacher, insofar as his teaching was restricted to what he had received as a student, he will “teach as a vicar, or image of [their] teacher” without his teacher having lost his or her essential knowledge (p. 135). The people retain their right despite the transmission of authority, but they are not themselves the teacher—political authority is not to be conceived along utilitarian lines as merely instrumental to the desire-satisfaction of atomized individuals. It is the people as a whole who transfer their right, and political authority takes on the responsibility of directing members of the people to the common good of the community that they share. The transmission principle cannot be reduced to the consent of individual members of society.

Unfortunately, Aquinas never explores the problem of transmission with anything like the fullness of scope that it demands. Nevertheless, he does deal with a relevantly similar concern that might, I argue, shed further light on the transmission of authority. I have in mind his treatment of the transmission of original sin to the descendants of Adam. Aquinas claims that the transmission itself is a matter of faith, and so it cannot be rejected. His problem is to offer a reasonable explanation in the face of contending possibilities: whereas some argue that sin is transmitted through the rational soul, others hold that it is through the body that the sin is passed on to future generations. Even though both views are reasonable explanations for the transmission of defects in general, Aquinas rejects both views because neither is capable of accounting for the “guilt” that attaches to the transmission of original sin: “granted that the rational soul were transmitted, from the very fact that the stain on the child’s soul is not in its will, it would cease to be a guilty stain binding its subject to punishment”

²⁶ Simon objects to this “cab-driver” theory of political authority effectively in *Philosophy of Democratic Government* (1966), pp. 146-54.

LYCEUM

(*ST I-II 81, 1*). Aquinas wants an account that attaches blame to future generations for something that they themselves are not directly guilty of insofar as they appear not to be responsible for the actions of a predecessor. Even though any given human person, from time t_2 through t_n , did not will or will not have willed what was willed by Adam at t_1 , original sin requires that blame attach to each. Aquinas' solution runs as follows: taken individually no single person can be held responsible for the sin of their first parent, but since all future generations make up one body, a universal community of humanity, with obligations that stretch backward and forward in time, individuals can be held responsible for original sin *as parts of that whole*. Aquinas draws an analogy between the parts of the body and the intention of the agent: the hand that commits murder is not held guilty *as part* but is held guilty of the act *as part of the whole*. Aquinas himself makes the extension of his analogy to the political community: "even as in civil matters, all who are members of one community are reputed as one body, and the whole community as one man" (*ST I-II 81, 1*).

I want to extend the argument to what I take to be an analogous situation in the political sphere with respect to the transmission principle. If the transmission principle were reducible to several acts of consent by members of a political community, there would be no problem with accounting for how authority is transferred, though there would be real difficulties disambiguating the consensus view from consent theories in the social contract tradition. Essential to consent theories of political obligation are the transactions undertaken by individual members of the state.²⁷ Given the obvious privileging of individualism, voluntariness, and autonomy in such accounts, it is easy enough to account for their appeal: "Voluntarists claim that our political obligations can arise only from our voluntary choices to subject ourselves to the political authority of

²⁷ For a complete exposition of consent theory, see the following works by A. John Simmons, *Moral Principles and Political Obligations* (1979), especially chapters 3-4; *On the edge of anarchy/Locke, Consent, and the Limits of Society* (1993); *Justification and Legitimacy* (2001), essays 1-8; *Political Philosophy* (2008), chapter 3; and, with Christopher Heath Wellman, *Is there a duty to obey the law?* (2005), chapters 5-8.

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

others or to participate in the ongoing cooperative schemes of political life” (Simmons, 2001, p. 66). Voluntarists start from a position in which human persons are free and equal in an imagined condition (think Locke’s ‘state of nature’) of pre-political association. In the case of naturalized citizens, there is even solid evidence for such acts of self-obligation, though these can only account for a very limited number of the citizens of a state, and so such theories face extension problems. How is it that *all* are obligated if only some have assumed such obligations voluntarily? Alternative mechanisms must be found, such as the Lockean notion of tacit consent. Even that notion, however, faces an inability to account for the obligations of future generations.

Despite its shortcomings, modern consent accounts of political obligation presuppose that the only valid way in which an individual can be bound to obey the commands of authority is through their own act of consent. Whenever the transmission principle is construed in such a way as to be consistent with such accounts, it will not only fall afoul of the objections to consent theory, but also fail to represent Aquinas’ account of political authority. Aquinas is not a consent theorist. The duty to obey does not derive from an individual act of consent, and there is no need to interpret the transmission principle to entail a consent theory.

Maritain often speaks of consent in relation to the transmission of authority, but he is usually careful to insist that it is the consent of the *whole people* that he has in mind. However, I part company with him in the use of consent in place of transmission because of its connotations. Consider a few such cases:

The leaders of the people receive this right from the creative and conservative principle of nature through the channels of nature itself, that is, through the consent or will of the people or of the body of the community, through which authority always passes before being invested in the leaders. (2011, p. 30)

LYCEUM

What has been gained for the secular conscience...is the conviction that authority, or the right to exercise power, is held by the rulers of the earthly community only because the common consent has been manifested in them and because they have received their trust from the people. (2011, p. 31)

[T]he organs of government are then regarded...as having the source of their authority in God...yet as not taking on even by participation a sacred character. Once the organs are designated, authority resides in them, but in virtue of a certain *consensus*, of a free and vital determination made by the people, whose personification and vicar they are... This consent itself must be understood in various senses. It can be formulated or unformulated. In the system of hereditary monarchy, it is once for all given for an indeterminate future, both as to the form of the regime and the eventual holders of power. In the democratic regime, it is once for all given for an indeterminate future as regards the form of the regime, but it is periodically renewable as regards the holders of power. (1996, p. 278)

Although it derives from it by a long progress of degradation which goes from Althusius and Grotius to Rousseau, this myth of the contract is quite different from the *consensus* which the ancients allowed to have been at the beginning of human societies... The Rousseauist contract has its first cause in the deliberate will of man, not in nature, and it gives birth to a product of human art. (1970, p. 133)

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

In each of the passages cited, Maritain returns to the themes of naturalism and the divine origin of political authority; in each case he refers consent to the *whole people*, though in case (c) the whole people refers both to an original moment when a democracy is founded and to the on-going exercise of democratic participation through suffrage. The latter comes very close, I think, to reducing transmission to so many individual acts, unless we assume that majoritarianism is equivalent to the whole people. I see no reason for taking this to be the case in the thought of Aquinas, and I find the supposition difficult to understand on its own. If transmission is conceived of as consent, the search for adequate evidence of consent may prove interminable and will likely prove unsuccessful. If transmission, by contrast, is conceived of in a manner analogous to the passage of original sin from generation to generation, this search becomes otiose. It is safer, therefore, to construe transmission in Aquinas (and in Maritain too) as not requiring a specific voluntary act even though future generations are bound by the act of a previous whole people in founding the state they reside in, keeping in mind that the transmission is a vicarious one from a people who itself is the recipient of such power.

Since political obligations follow from a teleological-eudaimonist structure of human nature, Aristotelian naturalism in which nature with its final ends is given by the Creator, and the Pauline principle, they are decidedly non-voluntary and general. The duty to obey, therefore, is best understood in a manner analogous to Aquinas' account of original sin. It requires no recourse to mechanisms like tacit consent; rather, as parts of political community whose existence is extended over multiple generations, individuals are understood to have transmitted the authority placed in the whole people by God *as a single body* to successive governments or rulers. No specific individual person has transmitted this authority for two reasons: first, it is not theirs to transmit as individuals, and, second, since they are not to be thought of as individuals in relation to the political community but *as parts of the whole*, they have transmitted authority and continue to transmit authority just as future generations

LYCEUM

of subjects are recipients of the blame attached to original sin without having acted as individuals in the commission of the sinful act at time t_1 .

2. Finnis' Rejection of the Consensus View

In *Natural Law and Natural Rights*, John Finnis' analysis of political authority is rooted in a reconstruction of Aristotle's examination of communities in the *Politics*, from conjugal to political. Marriage and friendships are, according to Finnis' analysis, types of community that are integral to the well-being of individuals. Each type of community enables its members to share in common projects, to which or through which their own life-plans are dedicated or realized. Smaller communities satisfy specific tasks or functions, and any attempt by larger communities to appropriate these will violate what Finnis calls the principle of subsidiarity. Political community is distinguished from smaller communities primarily in terms of completeness: "an all-round association in which would be co-ordinated the initiatives and activities of individuals, of families, and of the vast network of intermediate associations" (1980, p. 147).²⁸ Although more complete, the larger community is constrained by the principle of subsidiarity in such a way that it should not seek to replace smaller communities. Instead, it has as one of its principal tasks or functions the coordination of individual interests and the tasks performed by the smaller, incomplete communities out of which it is composed. Despite his indebtedness to Aristotle's analysis, Finnis faults Aristotle for making "a premature generalization from incomplete empirical data" (p. 148)—the premature generalization in question is the identification of the Greek *polis* as the most complete form of community. Finnis believes the descriptor 'completeness' is as appropriate to, if not more so,

²⁸ Finnis' allegedly Aristotelian-inspired conception of the political community sounds much more modern and individualist than one might expect: "The point of this all-round association would be to secure the whole ensemble of material and other conditions including forms of collaboration, that tend to favour, facilitate, and foster the realization by each individual of his or her personal development" (p. 147).

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

the modern “territorial state” characteristic of the contemporary international system.

Finnis also believes that the claims made by states to completeness can be extended to one of the more important social institutions of modern states, the legal system. References to “law” or “the law” normally signify—Finnis speaks of this marking the “central case” of their application—the same complete community as the modern state. The completeness of the legal system is measurable by its scope, supremacy, and comprehensiveness. Legal systems claim to be seamless in the sense of encompassing all aspects of life within their jurisdiction; they also claim to be “the supreme authority for their respective community;” and they tend to flesh themselves out by co-opting rules from smaller associations (p. 148).

The self-image of the legal system has its “foundation, from the viewpoint of practical reasonableness, in the requirement that the activities of individuals, families, and specialized associations be co-ordinated” (p. 149). Let me elaborate on what Finnis means by the “viewpoint of practical reasonableness”. Finnis is a new natural law theorist who grounds his normative theory—moral, political, and legal—in a conception of the natural law. Finnis’ particular conception denies the traditional assumption of natural law theory, which holds that normative requirements are derived from human nature. For Finnis the primary precept of the natural law is underived. It provides a grounding for what he describes as a list of primary goods that are irreducible and incommensurable. These goods are the constitutive components of individual flourishing, though different individuals in different times and places will have different, and one must assume, opposed ordinal rankings of the goods that will constitute their specific flourishing condition. The political and legal systems function in a largely instrumental fashion, providing individuals with the material resources and economic and political stability necessary for the pursuit of the primary goods.

LYCEUM

Given the instrumental role attributed to the state and its legal system, it makes sense to attribute to the law the role of coordinating the various activities of associations and individuals within a state. Finnis claims that the need for the law to fulfil this function “derives partly from the requirements of impartiality between persons, and of impartiality as between the basic values and openness to all of them” (p. 149). Finnis is less clear about the justification of the coordinative function that he attributes to the legal system than I would like. One wants to know, for example, what makes the law particularly well-suited to meet the requirements of impartiality, whether between persons or basic values. It may be that Finnis merely means to indicate that the rule of law—the law’s application to all equally—by definition makes it an ideal candidate to satisfy these requirements, though that would seem more overtly true of impartiality between persons than between basic goods; and yet, even if the law is by definition particularly well-suited to meet these requirements, that will not justify the presupposition that the state and its legal system are primarily, if not exclusively, instrumental.

However, the instrumental character of the state seems to be a pre-condition not only for the requirements of impartiality but also for the coordinative requirement. A full defense of the coordinative function of the law will require a prior defense, therefore, of the instrumental character of political community. Finally, the very notion of derivation is left underdetermined. From the facts that all human beings are mortal and that Socrates is a human being, I can derive the conclusion that Socrates is mortal. What is the connection between the need for impartiality between or normative openness to the basic goods and the need for coordination. I might claim that the various activities of individuals and groups in a specified territory must be coordinated to avoid the inefficient duplication of services. Having made that claim, I am not, however, in a position to conclude that there must be some overarching authority created in order to ensure such coordination. It might be possible for the individuals and agencies to voluntarily coordinate their activities, independently of an authoritative agency. Where self-regulation is impossible, it might still be

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

possible for the individuals and agencies to organize matters in a way that authority rotated among the members of the community without the further need for a distinctly political organization of their activity. Of course, Finnis does claim that the coordinative function of the political community is only “partly” derived in this manner.

Finnis does have a fuller account of the “need and justification” of authority in political communities. There are obvious causes of the need for authority in a complete community—including, “the stupidity and incompetence of its members, their infirmity of purpose and want of devotion to the group, their selfishness and malice, their readiness to exploit and to ‘free ride’”—but these accidental causes are not what Finnis has in mind, though he is aware that in most communities these will be good reasons for having political authority. He seeks to answer the following question instead: “In a community free from these vices, would authority be needed, or justified?” (p. 231). The aim is to provide an essential, as opposed to an accidental, justification of political authority. Notice that in the cases of some weakness, the authority is only a substitute for the failed capacity at issue. In a community where such incapacities were absent, the justification of authority that depends on them would also be absent. Finnis has something more permanent in mind, though what precisely he is after will be harder to discover than one might expect.

Finnis suggests that wisdom—the perfection of the intellectual powers— itself can bring it about that the political community needs authority: “the greater the intelligence and skill of a group’s members, and the greater their commitment and dedication to common purposes and common good, the *more* authority and regulation may be required, to enable that group to achieve its common purpose, common good” (p. 231). I do not see why a greater degree of intelligence among the governed would call for *more* authority than in cases where stupidity was the norm, but Finnis’ case doesn’t rest on this claim and I think it best to ignore it. The principal idea for Finnis is that intelligence will increase the number of truly good options faced by a particular community, but

LYCEUM

the fact that there are several *equally good options* for the community as a whole to pursue requires some decision-making procedure to ensure that some *one* decision be made. Finnis then examines the following two scenarios:

S₁: A community must decide between two options, a good one, *G*, and a bad one, *B*; assuming that everyone in the community is intelligent, they will unanimously choose *G*, and there will be no need for an authority.

S₂: The same community, composed of intelligent citizens, must choose between several good options {*G*₁, *G*₂, *G*₃...*G*₁₀}; since all of the options are, all things considered, equally good, though mutually exclusive alternatives, the community must devise some means of arriving at a decision. Unanimity is unavailable to them, but not as a result of any deficiency of any of the members. They are all intelligent citizens. There is also no deficiency in any of the goods: each good is a genuine good, but, by hypothesis, only one good from the series can be chosen. Each of the members of the community could defend one or several of the alternatives, but not because of any shortcoming or deficiency in either the members of the community or the series of goods from which they are required to choose. In such a situation, authority has the role of providing a mechanism by means of which a determination is reached.

These scenarios exhaust the possible alternatives open to communities faced with making decisions that involve coordinating the activities and interests of several groups and individuals. For the most part, unanimity is not a real option because in most cases, and this would be especially true of a community with an

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

intelligent citizenry, the options available involve “reasonable, and appropriate solutions, none of which, however, would amount to a solution unless adopted to the exclusion of the other solutions available, reasonable, and appropriate for that problem” (p. 232). In the absence of unanimity, and assuming that there are really only two alternatives, then there is a need for political authority. This need is not contingent on any deficiency, either in the means between which a choice must be made or in the members responsible for making the choice.

What are the central objections to Finnis' coordination account of political authority? I want to suggest that, apart from internal shortcomings, Finnis' theory takes for granted, but is insufficiently inarticulate about its dependence on, a fairly robust or thick conception of Aristotelian naturalism, despite the fact that Finnis denies that political authority derives from human nature. That is to say, there would be no need for the coordination of social groups if human beings did not have an inclination that predisposed them to live in social groups. Finnis can deny the prior existence of such a disposition, opting for something like the state of nature, where coordination substitutes for the more chaotic social relations that obtain therein; but in that case, his account of political authority will move in the direction of consent theory, because persons in the uncoordinated state of nature will have to choose to depart therefrom once they are persuaded of the benefits of doing so.

In contrast to Finnis, Aquinas takes the fact that it is natural to human beings to live in social groups as a necessary starting point for his hybrid account of political authority. What the two principles of the consensus view add to that starting point is an account of the origin of the natural tendency or inclination that human beings have to live in social groups. Looking to the origin of that nature directs the argument toward a Creator of nature. This is, of course, the starting point for the Pauline principle; the transmission principle accounts for the existence of the right in political rulers, a right that is initially in and remains with the whole people.

LYCEUM

2.1 Finnis' Rejection of "Transmission" in *NLNR*

Finnis is silent on the Pauline principle in *Natural Law and Natural Rights*, but he has something to say about the second principle of the consensus view, the transmission principle. As has been mentioned, unlike those contemporary authors who simply neglect the consensus view, Finnis explicitly argues against the transmission principle. He dismisses the principal texts used in support of the transmission principle—that is, *ST* I-II 90, 3 and 97, *3ad3*—as “ambiguous and unsatisfactory remarks of Aquinas” (p. 257). Since these are primary texts on which the transmission principle is defended, it is unfortunate that Finnis does not provide his reasons for believing them to be ambiguous and unsatisfactory. I have pointed out that I too find the discussion of custom somewhat difficult to contextualize, but it doesn't seem to me to be so difficult as to be unclear in its leading idea that authority rests with the people. With the second passage, it would have been very helpful to have Finnis address Maritain's useful clarification of the concept of vicariousness, but Finnis doesn't mention this as far as I can tell.

Instead, the heart of his criticism takes up an argument proposed by Bellarmine and widely cited among defenders of the consensus view. Finnis treats Bellarmine's argument as an example of “theories of governmental legitimacy and political obligation which tacitly assume that the present authority of particular rulers must rest on some prior authority (of custom; or of the community over itself, granted away to the ruler by transmission or alienation...)” (p. 247). Such theories correctly, Finnis continues, hold that “*all* the members of a community are entitled in justice to a certain concern and respect” (pp. 247-8). While in a vague way this is true, it is clearly *not* one of the central premises of the transmission principle, which, recall, takes as its starting point the Pauline principle, and then argues that the authority that human beings exercise over one another is caused by the transmission of the authority that is predicated of the community as a whole. There is no mention in Aquinas or other

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

defenders of the consensus view of “concern and respect” as a central premise in the justification of political obligation.

Here is Finnis’ reconstruction of Bellarmine’s argument: “Natural reasonableness requires that there be governmental authority; But natural reasonableness does not identify any particular man or class as the bearer of governmental authority; Therefore natural reasonableness requires that the bearer of governmental authority be the multitude, the whole community itself” (p. 248). I shall argue below that this is a serious misrepresentation of Bellarmine and of the transmission principle as a whole. For now, however, let me rehearse Finnis’ criticism of the syllogism that he attributes to Bellarmine, since he not only describes Bellarmine’s version as “helpfully clear” but also as revelatory of “the fallacy in his theory, and in all such ‘transmission’ theories” (p. 248). Where does Bellarmine’s fallacy lie? There is nothing wrong, according to Finnis, with either of the two premises. The suggestion is that they are both true, but that they do not support the conclusion, which is “intrinsically implausible.” As a term of art I shall call this the “absence thesis”: in the absence of a particular ruler or rulers (the second premise of Finnis’ reconstruction), authority must fall to the people as a whole. Aquinas certainly does not hold the absence thesis, but there are only two key occasions in the work of Aquinas where he seems to claim simply that political authority belongs to the people. Does Bellarmine? He *does* but it does not have the impact on his argument that Finnis seems to think it does.

The first premise, without much wrenching, can be construed as wholly consistent with Finnis’ own account of the origin of political authority. Natural reasonableness requires the existence of governmental authority in order to identify a means of coordinating human activity. Equally, though not necessary to Finnis’ theory, the second premise is not opposed to that theory: Finnis doesn’t propose that the need for a solution to coordination problems helps to identify specific rulers, merely the need for the same.

LYCEUM

To understand Finnis' reasoning, therefore, we should keep in mind his defence of political authority as *the* solution to coordination problems in social and political orders. What Finnis argues is that the transmission principle, as he has presented it, violates this justification for political authority. If political authority derives exclusively from the need, as the only alternative to unanimity (recall that the two scenarios, S_1 and S_2 , were exhaustive), and in light of the rarity of achieving unanimity, for the "solution of practical co-ordination problems which involve or concern everyone in the community," then to say that authority is caused by the community is to assume either that "there is no authority in this community" or "it amounts to saying something else, by way of a confusing legal fiction or ideological manner of speaking, about the location of authority in *some* communities" (p. 248). In the first alternative, the solution of coordination problems will depend on either unanimity or force. In the second alternative, Finnis asks us to imagine that the problem will be solved by pretending that the participation of individuals in such imagined communities will be exercising authority through choosing the person or persons in whom it should be placed. The problem with this latter alternative is, according to Finnis, that for voters in the minority, the casting of a vote is not really an exercise of authority. I assume that he means that exercises of authority are only those that carry the day. To some extent, I want simply to argue that Finnis has put the cart before the horse. Keep in mind that he accepts both premises that he attributes to the transmission principle. The first premise, recall, required further justification—at least in the version that we find in Aquinas and the defenders of the consensus view. That further justification leads one to the cause of nature. The solution of coordination problems is wholly unnecessary in a world where human nature is not governed by the ordinating Law of God. Finnis simply takes sociability for granted as a dictate of "natural reasonableness" as if that did not in itself, apart from providing it with a source, reduce to an ideological way of speaking.

But there is a much more significant set of problems with Finnis' critique. The chief problem is that it simply gets the transmission principle wrong. I

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

realize that this is a bold statement, but anyone with even a limited familiarity with the consensus view would have to conclude the same, however uncomfortable it might make one feel in doing so. There are at least two reasons for believing that Finnis has fundamentally misstated the transmission principle. To begin with, in the syllogism he attributes to Bellarmine, as a representative defender of the consensus view, the premise that “natural reasonableness” is silent on the issue of who should rule. There is a sense in which this is true. All advocates of the consensus view would agree that the transmission principle does not concern who governs. But none of them, to my knowledge, use this as a premise in their defense of the position. The fact is that they explicitly exclude from consideration any and all talk of a distinct governing personnel *because* they are all convinced that this is a subordinate issue to and distinct concern from the foundation of authority itself. Indeed, it is precisely here that human art enters into political statesmanship. The consensus view treats the form of government as morally indifferent, and so it leaves that open to the prudential judgement of the people. What the consensus view seeks to establish is the origin and justification of political authority.

Given that defenders of the consensus view agree in excluding concern with specific regime types from their arguments, Finnis' attribution of such a premise to the transmission theory is perplexing. He has not only set up a straw person argument, but has done so by attributing to his opponents a premise to which they are in unanimous opposition. In addition, in Finnis' reconstruction he makes no mention of the Pauline principle. On my reading of the consensus view, however, the Pauline principle must be taken to be the first premise. In this case one might excuse Finnis because, as I have argued above, too often the consensus view is inarticulate in this regard, though Maritain stands out as an important exception. Still, I assume that some of the inarticulacy results from the fact that the defenders of the consensus view took for granted their shared starting point. On my reconstruction of the consensus view, the Aristotelian

LYCEUM

premise about human nature and its need for political community leads back to the Pauline principle.

I have charged Finnis with misstating the consensus view, and so I want to conclude this section by referring directly to the arguments of Bellarmine in its defense. It was Bellarmine's construction that Finnis faulted with arriving at a faulty conclusion from premises that Finnis was ready to accept. What is unclear, however, is that these *are* Bellarmine's premises. In his *On Laymen or Secular People*, Bellarmine has his sights set not only on the justification of political authority but also that of holy war. With respect to the former, he positions himself as a defender of the naturalism of political authority against Protestant reformers who conceived of political authority as dependent on God's grace. As might be expected, Bellarmine regularly returns to the Pauline text to support his view that authority has a lawful ground (2012, p. 10). What is essential to point out, however, is that Bellarmine does not offer just one justification. He argues in a variety of modes. For example, he begins by arguing from final causality: "political authority is so natural and necessary to humankind that it cannot be removed without destroying nature itself. In fact, by nature man is a social animal...[and so, unlike nonhuman animals] he absolutely cannot live by himself" (p. 18). Bellarmine proceeds to list the ways in which authority is necessitated by nature: food, shelter, clothing, defense, and education. Drawing directly on the *Politics*, he goes on to list speech as itself requiring human linguistic communities. Chief among the functions of authority so conceived will be directing subjects to the common good: "if human nature requires a social life, certainly it also requires a government and a ruler, as it is impossible that a multitude can last long unless there is somebody to hold it together and be in charge of the common good (p. 19). Human government is not only necessary but lawful—it is not to be conceived as a punishment for human sinfulness (p. 19).

It is only when Bellarmine turns from arguments connected to Aristotelian naturalism, that he introduces the Pauline principle explicitly, and without the Pauline principle there would be no authority from above to be transmitted, a

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

fact that Finnis remains silent about. Having treated of the origin and lawfulness of political authority from the point of view of final causality, Bellarmine turns to consider efficient causality: “as it is certain that political authority comes from God, from Whom nothing proceeds but the good and the lawful,” political authority must itself be good (p. 21). What reason or reasons does Bellarmine offer as evidence in support of this conclusion? He begins by making two crucial clarifications. First, he takes for granted that authority must be derived from human nature (his argument from final causality), and so he simply acknowledges, in a manner wholly unavailable to Aristotle, that God, as the source of that nature, is also the source of authority insofar as it is God who gives to each thing its nature. It is at this point in the argument, incidentally, that Bellarmine claims that he is only accounting for political authority in general, and as such is not required to reflect on particular forms of rule, such as monarchy, aristocracy and so forth. Second, Bellarmine rejects consent as an adequate ground for political authority: “this authority is of natural law, as it does not depend upon men’s consent. In fact, whether or not they want to, men must be ruled by somebody unless they want humankind to perish, which is against the inclination of nature” (p. 21). Bellarmine again recalls the Pauline text—this time claiming that the derivation of the natural law from the eternal is what was intended by the Apostle (p. 22).

At this stage in his argument, Bellarmine *does* introduce what I earlier called the “absence thesis.” Bellarmine writes: “this authority immediately resides in the entire multitude as its subject because this authority is of divine law. But divine law did not give this authority to any particular man; therefore it gave it to all” (p. 22). Let me add two observations before I return to the absence thesis. Bellarmine offers two other reasons for accepting that authority is given by the eternal law to the people as a whole, and so, even if Finnis were successful in dismissing the absence thesis, it would not suffice as a critique of the consensus view as presented by Bellarmine. On the one hand, Bellarmine argues from natural equality: outside of positive law all human beings are

LYCEUM

created equal, and so there is no reason to think that any particular person or group of persons should rule. On the other hand, the political community is a perfect community, and, as such, it must have the power to defend itself from demise. This is the type of power exercised by political authority through internal policing and national defense (p. 22).

Finnis objects to the absence thesis that it would require the people to transmit their authority, but such a collective action can only occur through unanimity or authoritative coordination. In that case, the need for coordination precedes any alleged transmission, rendering it unnecessary in fact because coordination would be the only solution to transmission. I have already argued that Finnis' coordination account itself runs up against a priority objection: some account of the need for social and political life must be presupposed by Finnis in order for there to be a need for coordination. With respect to his remarks on Bellarmine, I think they assume a consent framework: the absence thesis is objectionable because it creates the either unanimity or coordination disjunct. If the people as a whole transmit their authority to the political authority, then it avoids the dilemma to which Finnis objects. And there is good reason for thinking that Bellarmine is uninterested in consent. He explicitly asserts that the people must transfer their authority as a matter of the law of nature: "the commonwealth is obliged to transfer it to one or a few" (p. 22). Consent does have a role to play, but where one would expect it: at the level of the *ius gentium*. The choice of a specific regime, which is not of the natural law, falls under the type of positive law known as the law of nations (p. 22). It is at this level that "human deliberation and decision" play a role, but that is true of positive law more generally in its distinction from natural law. This helps to explain the assumption of legitimacy by conquerors: conquered peoples who eventually accept the rule of particular regimes, Bellarmine gives the example of the unjust occupation of Gaul by the Franks, can be said to have consented to the rule. Yet this is a principle not of natural law but of the law of nations (p. 25).

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

2.2 Finnis' Rejection of Both Principles of the Consensus View in Aquinas

In *Natural Law and Natural Rights*, Finnis need not have tackled the Pauline principle, since the work was not intended as an exegetical account of the political and social theory of Aquinas. Indeed, in a limited way, Finnis attempted to steer his argument more in the direction of an Aristotelian account than one drawn from Aquinas' natural law theory. However, in his *Aquinas* Finnis firmly grounds his theory in the work of Aquinas. He not only returns to the transmission principle, which he again rejects, but also takes up the Pauline principle, though this fares little better.

Finnis complains that "Aquinas' position is clearer in its articulation than it is substance," when it comes to the account of political authority's legislative power. It is at this juncture that Finnis glances at *ST* I-II 97, 3ad3, the text in which Aquinas considers the power of the people to establish laws through custom. This is one of the texts used by advocates of the transmission principle to demonstrate its derivation from the works of Aquinas. If Aquinas is unclear in terms of the substance in this instance—and I don't see that he is—he is clear enough elsewhere about the legislative authority that persons or institutions that have care for the community hold: "Whoever are subject to a king, are bound to observe his law which he makes for all in general" (I-II 98, 5ad1).

Finnis' discussion of the two principles of the consensus view in this work comes during his discussion of what he takes to be Aquinas' defense of limited government. Finnis argues that the right allowed by Aquinas of subjects to disobey their rulers entails limited government (p. 258). While this sounds right, it risks overstating Aquinas' permission: individual prescriptions can be rejected, but that doesn't void the authority that made them, and, more importantly, insofar as they are dictates of authority the authority itself still carries some weight even in non-binding (because unjust) laws. I mention the context because it helps shed light on Finnis' overall presentation of Aquinas. Finnis will, for example, speak of an "implicit social contract" in spelling out Aquinas' political

LYCEUM

theory (p. 261), and, more surprisingly, of the duty owed to political authority as one “strictly speaking, [owed] to the rulers themselves but rather to, if anyone, their fellow citizens” (p. 264). In fact this is the point of saying that political authority has care of the community: “To extend this idea of representation into an idea that authority must, or should, have been transmitted by some procedure of transference (however implicit or tacit) from the people to their representative(s) is to miss the point of the construction, and to convert it into a fiction or a sometimes inappropriate requirement for just government” (p. 264). So much for the transmission theory. Finnis simply repeats the claim that the transmission principle leads to an unacceptable dilemma (pp. 264-65). I have already suggested why this argument is unconvincing. But there are further risks associated with Finnis’ appropriation. He continues, “if the people have the right to make provisions regarding their king or other presiding ruler, they can properly curb or wholly remove the authority” (p. 265). How does such a picture avoid being reduced to a consent account? Does not Finnis make the so-called “implicit social contract” that he detects in Aquinas almost the equivalent of full-blown social contract, wherein governments are limited by provisions intended to protect the parties to the original agreement?

Conclusion

As a further sign of the displacement of the consensus view on the origin of political authority, a recent collection of essays dedicated to Aquinas’ commentary on Paul’s letter to the Romans not only lacks a contribution on chapter 13, despite the significance for Aquinas of that chapter, but also lacks any mention of political authority and obligation in the index of the work. Indeed, in one essay where chapter 13 is mentioned, the author, John Boyle, whose essay is devoted to showing how the arguments of the *Summa* can be taken to illuminate those of the commentary, goes out of his way to displace, downplay, and minimize the important interpenetration of the texts in this one case:

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

At the same time, the division of the text imposes its own constraints on the interpretation of the text, as can be seen in St. Thomas's interpretation of the first ten verses of chapter 13, which have been the subject of much deliberation on the relation of the Christian to the political order. While that relation is part of St. Thomas's consideration, it is not at the heart of it. The issue is the virtue of justice lived in relation both to superiors and to fellow men. The unity of the division stipulates the higher context of this justice, namely, the perfection of man, a perfection here horizontally situated in relation to man's holiness and purity. The relation to the political order is thus rendered secondary, or perhaps better, reconfigured in relation to the spiritual perfection of man in grace. (Boyle, 2012, p. 79)

To say that political authority is not "at the heart of" this passage or that it is "rendered secondary" or "reconfigured" is simply to ignore the role played by the passage throughout Aquinas' *corpus*. For Aquinas, political authority is an *integral* part of the "higher context" of any conception of justice, despite the discomfort that this appears to cause contemporary scholars. Human perfection is hierarchically situated in such a way that political authority is itself an indispensable component. As I have shown, Aquinas takes *Romans* 13: 1-2 as the principal New Testament support not only for the correlation of political authority and political obligation, but also for the realization of the human good. It *is* the foundation upon which he builds his account of human government insofar as the morally obligatory character of the political authority derives from and depends upon the Creator of human nature. Aquinas marries his Aristotelian naturalism—his theory that authority is, as it were, built-in to human nature inasmuch as human flourishing depends upon political community—to the

LYCEUM

supernatural cause of that nature, and in doing so regularly avails himself of the Pauline text as support for this hybrid account of political authority.

Aquinas' interpretation of and reliance on the Pauline principle, along with the concomitant transmission principle, in his account of political authority has a remarkable pedigree. In *Diuturnum*, Leo XIII refers directly to the Pauline principle as being at the center of a proper understanding of political authority, and he traces its development through Augustine, St. John Chrysostom, St. Gregory the Great, and Aquinas (§10). Even though he speaks against the prevailing opinions, which "reject, with more boldness than formerly, every restraint of authority," Leo defends the necessity of political authority as susceptible of rational and religious demonstration, especially against the dominance of the social contract tradition:

Those who believe civil society to have risen from the free consent of men, looking for the origin of its authority from the same source, say that each individual has given up something of his right, and that voluntarily every person has put himself into the power of the one man in whose person the whole of those rights has been centred. But it is a great error not to see, what is manifest, that men, as they are not a nomad race, have been created, without their own free will, for a natural community of life. It is plain, moreover, that the pact which they allege is openly a falsehood and a fiction, and that it has no authority to confer on political power such great force, dignity, and firmness as the safety of the State and the common good of the citizens require. Then only will the government have all those ornaments and guarantees, when it is understood to emanate from God as its august and most sacred source. (§12)

The Consensus View's Two Principles of Political Authority: Pauline and Transmission

On the one hand, there is a clear appeal to Aristotelian naturalism, which is repeated elsewhere in the encyclical: “nature... wills that man should live in a civil society; and this is clearly shown both by the faculty of language... and by numerous innate desires of the mind... which men isolated cannot procure, but which they can procure when joined and associated with others” (§11). To do so, however, there must exist “one to govern the wills of individuals, in such a way as to make, as it were, one will out of many, and to impel them rightly and orderly to the common good.” On the other hand, there is an equally clear appeal to the Pauline principle: “no man has in himself or of himself the power of constraining the free will of others by fetters of authority of this kind. This power resides solely in God, the Creator and Legislator of all things; and it is necessary that those who exercise it should do it as having received it from God.”²⁹ Of course, it would be easy to object that Leo XIII, as one who called for a return to the works of Aquinas, was himself responsible for the consensus view’s agreement on these key principles. It would be more important, therefore, to cite a more contemporary authority to suggest where Catholic political thought stands today in relation to the consensus view. I can think of no better authority than the *Catechism* for this purpose: “Every human community needs an authority to govern it. The foundation of such authority lies in human nature. It is necessary for the unity of the state” and “The authority required by the moral order derives from God” (1997, §§1898-99).³⁰ The Thomistic synthesis of

²⁹ Leo refers to the key Pauline passage from *Romans* twice (§9 & §14).

³⁰ The warrant for the principle is, not surprisingly, the Pauline text which is quoted immediately after the passage cited above and is itself the sole authority quoted in the summary at §1918. Aristotelian naturalism and the Pauline principle also inform the discussion of the political community in *Gaudium et Spes* (1988, §§ 73-75). These themes are repeated again more recently in the *Compendium of the Social Doctrine of the Church*, which includes an unmistakable reference also to the transmission principle of the consensus view: “In various forms, the people transfers the exercise of sovereignty to those whom it freely elects as its representatives, but it preserves the prerogative to assert this sovereignty in evaluating the work of those charged with governing and also in replacing them when they do not fulfil their functions satisfactorily” (§395).

LYCEUM

Aristotelian naturalism with the two principles of the consensus view continue to guide the teachings of the Church on the foundation of political authority.

References

- Aquinas, T. (2012). *Commentary on the Letter of Saint Paul to the Romans*. (F.R. Larcher, Trans.). Lander, WY: The Aquinas Institute for the Study of Sacred Doctrine. (Original work published, 1272-73).
- Aquinas, T. (2006). *Commentary on the Epistle to the Hebrews*. (C. Baer, Trans. and Ed.). South Bend, IN: St. Augustine's Press. (Original work published, 1272-73).
- Aquinas, T. (1985). *Commentary on the Sentences of Peter Lombard*. In *Civil Authority in Medieval Philosophy/ Lombard, Aquinas and Bonaventure* (M. Malloy, Trans. and Ed.). Washington, DC: University Press of America. (Original work published, 1254-56).
- Aquinas, T. (1956). *Summa Contra Gentiles*. (V. Bourke, Trans.). South Bend, IN: University of Notre Dame Press. (Original work published, 1259-65).
- Aquinas, T. (1948). *Summa Theologica*. (Fathers of the English Dominican Province, Trans.). New York, NY: Benzinger Brothers. (Original work published, 1268-73).
- Aquinas, T. (1902). *An Apology for the Religious Orders*. (J. Procter, Trans.). London, UK: Sands. (Original work published, 1271).
- Augustine. (2000). *The City of God*. (M. Dods, Trans.). New York, NY: The Modern Library. (Original work published, 426).

***The Consensus View's Two Principles of Political Authority:
Pauline and Transmission***

- Baur, Michael. (2012). Law and Natural Law. In B. Davies and E. Stump (Eds.). *The Oxford Handbook of Aquinas* (238-54). New York, NY: Oxford University Press.
- Bellarmino, Robert. (2012). *On Temporal and Spiritual Authority/On Laymen or Secular People, On the Temporal Power of the Pope Against William Barclay, and On the Primary Duty of the Supreme Pontiff*. (S. Tutino, Ed. and Trans.). Indianapolis, IN: Liberty Fund.
- Black, A. (1992). *Political Thought in Europe 1250-1450*. New York, NY: Cambridge University Press.
- Blythe, J. (1986). The mixed constitution and the distinction between regal and political power in the work of Thomas Aquinas. *Journal of the History of Ideas* 47, (4 Oct-Dec), 547-65.
- Boyle, J. (2012). On the relation of St. Thomas's Commentary on the *Romans* to the *Summa theologiae*. In M. Levering and M. Dauphinais (Eds.), *Reading Romans with St. Thomas Aquinas* (75-82). Washington, DC: Catholic University of America Press.
- Catechism of the Catholic Church*. (2nd Ed.). (2000). Vatican City, VA: Libreria Editrice.
- Deane, H. (1963). *The Political and Social Ideas of St. Augustine*. New York, NY: Columbia University Press.
- Finnis, J. (1980). *Natural Law and Natural Rights*. Clarendon Law Series. New York, NY: Oxford University Press.

LYCEUM

- Finnis, J. (2004). *Aquinas/ Moral, Political and Legal Theory*. Founders of Modern Political and Social Thought. New York, NY: Oxford University Press. (Original work published, 1994).
- Hibbs, T. (2001). *Virtue's Splendor: Wisdom, Prudence, and the Human Good*. New York, NY: Fordham University Press.
- Hittinger, J. (2002). *Liberty, Wisdom, and Grace: Thomism and Democratic Political Theory*. Lanham, MD: Lexington Books.
- Konyndyk DeYoung, R., McCluskey, C., and Van Dyke, C. (2009). *Aquinas's Ethics/Metaphysical Foundations, Moral Theory, and Theological Context*. South Bend, IN: University of Notre Dame Press.
- Leo XIII. (1954). *Diuturnum*. In E. Gilson (Ed.), *The Church Speaks to the Modern World/ The Social Teachings of Leo XIII*, (140-56). New York, NY: Image Books. (Original work published, 1881).
- Luscombe, D. E. (1982). The state of nature and the origin of the state. In N. Kretzmann, A. Kenny, J. Pinborg, and E. Stump (Eds.). *The Cambridge History of Later Medieval Philosophy (757-770)*. New York, NY: Cambridge University Press.
- MacGuigan, M. (1957). *The Best Form of Government in the Philosophy of St. Thomas Aquinas*. Toronto, CA: unpublished PhD dissertation, University of Toronto.
- MacIntyre, A. (1988). *Whose Justice? Which Rationality?* South Bend, IN: University of Notre Dame Press.

***The Consensus View's Two Principles of Political Authority:
Pauline and Transmission***

- MacIntyre, A. (1990). *Three Rival Versions of Moral Enquiry/ Encyclopedia, Genealogy, and Tradition*. South Bend, IN: University of Notre Dame Press.
- Maritain, J. (2011). *Christianity and Democracy and The Rights of Man and the Natural Law*. (D. Anson, Trans.). San Francisco, CA: Ignatius Press. (Original work published, 1942 and 1943).
- Maritain, J. (1996). *Integral Humanism*. In Otto Bird (Ed.) *The Collected Works of Jacques Maritain* vol. IX. (O. Bird and J. Evans, Trans.). South Bend, IN: University of Notre Dame Press. (Original work published, 1936).
- Maritain, J. (1972a). The Catholic Church and Social Progress. In *Ransoming the Time*. (H. Lorin Binsse, Trans.). New York, NY: Gordian Press. (Original work published, 1939).
- Maritain, J. (1972b). The Political Ideas of Pascal. In *Ransoming the Time*. (H. Lorin Binsse, Trans.). New York, NY: Gordian Press. (Original work published, 1941).
- Maritain, J. (1970). *Three Reformers: Luther, Descartes, and Rousseau*. New York, NY: Thomas Y. Crowell Company. (Original work published, 1925).
- Maritain, J. (1951). *Man and the State*. Chicago, IL: University of Chicago Press.
- Maritain, J. (1940). *Scholasticism and Politics*. (M. Adler, Trans.). New York, NY: The Macmillan Company.

LYCEUM

- McCoy, C. (1989). The problem of the origin of political authority. In *On the Intelligibility of Political Philosophy*. J. Schall and J. Schrems (Eds.). Washington, DC: Catholic University of America Press.
- Murphy, M. (2009). *Natural Law in Jurisprudence and Politics*. New York, NY: Cambridge University Press.
- Parsons, W. (1939). *Which Way, Democracy?* New York, NY: The Macmillan Company.
- Parsons, W. (1940a). The Influence of *Romans XIII* on Pre-Augustinian Christian Political Thought. *Theological Studies*, 1 (4), 337-64.
- Parsons, W. (1940b). The Influence of *Romans XIII* on Christian Political Thought II. Augustine to Hincmar. *Theological Studies*, 2 (3), 325-46.
- Parsons, W. (1941). St. Thomas Aquinas and Popular Sovereignty. *Thought: Fordham University Quarterly*, 16 (September), 473-92.
- Pasnau, R. and Shields, C. (2004). *The Philosophy of Aquinas*. Boulder, CO: Westview Press.
- Pontifical Council for Justice and Peace. (2005). *Compendium of the Social Doctrine of the Church*. Vatican City, VA: Editrice Vaticana.
- Porter, J. (2010). *Ministers of the Law/ A Natural Law Theory of Legal Authority*. Grand Rapids, MI: Wm. B. Eerdmans Publishing Company.
- Rommen, H. (1945). *The State in Catholic Thought/A Treatise in Political Philosophy*. St. Louis, MO: B. Herder Book Company.

***The Consensus View's Two Principles of Political Authority:
Pauline and Transmission***

- Sigmund, P. (1993). Law and Politics. In N. Kretzmann and E. Stump (Eds.). *The Cambridge Companion to Aquinas* (217-31). New York, NY: Cambridge University Press.
- Simmons, A. J. (2008). *Political Philosophy*. New York, NY: Oxford University Press.
- Simmons, A. J. (2001). *Justification and Legitimacy*. New York, NY: Cambridge University Press.
- Simmons, A. J. (1993). *On the Edge of Anarchy/ Locke, Consent, and the Limits of Society*. Princeton, NJ: Princeton University Press.
- Simmons, A. J. (1979). *Moral Principles and Political Obligations*. Princeton, NJ: Princeton University Press.
- Simmons, A. J. and C. Wellman. (2005). *Is there a duty to obey the law?* New York, NY: Cambridge University Press.
- Simon, Y. (1940). *Nature and Functions of Authority*. Marquette, WI: Marquette University Press.
- Simon, Y. (1966). *Philosophy of Democratic Government*. Chicago, IL: University of Chicago Press.
- Simon, Y. (1968). *Freedom and Community*. New York, NY: Fordham University Press.
- Simon, Y. (1980). *A General Theory of Authority*. South Bend, IN: University of Notre Dame Press.

LYCEUM

Vatican II. (1988). *Gaudium et spes*. In A. Flannery (Ed.), *Vatican II: The Conciliar and Post Conciliar Documents*, vol. I (903-1014). Northport, NY: Costello Publishing Company. (Original work published, 1965).

Extended Mind and the Extension of a Self

Sangkyu Shin

I.

Digital technology has made various forms of convergence possible that would not otherwise have been possible, such as the convergence of hardware devices, media, and various applications. The most interesting form of convergence, however, is that between digital devices, such as the iPhone, and their human users. If one considers a smartphone an externalization of one's own cognitive processes, there is a convergence between subject and instrument, between one's cognitive subjectivity and a device that assists or substitutes for one's cognitive activities. The convergence of human beings and digital technology requires reexamining the boundaries of humans, instruments, and environments in a fundamental way. As the boundaries between self and instrument become vague and unclear, a new approach to the self becomes necessary.

Recently, there have been several attempts in philosophy and psychology to understand the human mind as extending into the external world. For example, Andy Clark and David Chalmers (1998) claim that when we perform cognitive activities with the help of external tools, the extra-cranial states or processes involved in those tools may be regarded as parts of our mind or our cognitive processes. The so-called "extended mind thesis" by Clark and Chalmers in their article "The Extended Mind" contends that our mind is not confined within an organism (inside our body) or our brain, but it can be extended outward over the boundary of our skin (pp. 7-19). In the foreword to Andy Clark's *Supersizing the Mind*, for example, Chalmers (2008) claims, "My iPhone is not my tool, or at least it is not wholly my tool. Parts of it have become parts of me" (p. x). This amounts to the claim that if external devices such as iPhones are coupled with our brains in an appropriate way, such devices themselves can be considered parts of our minds.

LYCEUM

What exactly is being extended when it is claimed that our mind extends beyond our body? According to Clark and Chalmers (1998), the extension of our mind can be understood in three different ways. The first is the extension of cognitive *processes*. Clark and Chalmers take an example for this from two ways of conducting a Tetris game. In solving a Tetris puzzle, we may do it while rotating each Tetris piece in the mind (or in the brain). On the other hand, we may literally rotate the piece on the screen by pressing a button on the keyboard while checking its congruence with our eyes. The latter is a case of the cognitive process occurring externally.

The second kind is the extension of cognitive *states*, such as dispositional beliefs or memories. To illustrate this claim, they give an example of the sentences or information stored in a notebook kept by Otto, who is suffering from Alzheimer's disease. There is a sentence recording a museum's address: "The museum is located on 53rd Street." Whenever Otto goes to the museum, he opens his notebook, checks the address, and uses it. The information stored in the notebook falls under a dispositional belief, the second kind of extension.

These extensions of cognitive processes and states are extensions of the mind in a narrow sense. The third kind of extension is the extension of a *self*, i.e., the extension of an agent or a person. An individual can compose a transitory, integrated system when coupled with external resources, and Clark and Chalmers (1998) claim that this can be regarded as an extended self:

Does the extended mind imply an extended self? It seems so The information in Otto's notebook, for example, is a central part of his identity as a cognitive agent. What this comes to is that Otto himself is best regarded as an extended system, a coupling of biological organism and external resources. To consistently resist this conclusion, we would have to shrink the self into a mere bundle of occurrent states, severely threatening its deep psychological continuity. Far better to take the broader

Extended Mind and the Extension of a Self

view, and see agents themselves as spread into the world.
(p. 18)

We may feel uncomfortable with the thesis of the extended mind in several ways. Among these ways, even if we grant a sort of cognitive character to the external states or process, the most profound discomfort is probably related to a question about the extended self, namely how it can be considered the extension of “my” mind. In this paper, I intend to advocate one way of understanding the extended self by answering this question. For this purpose, I will briefly discuss and criticize two recent criticisms by Eric T. Olson and Lynne Rudder Baker against the idea of the extended self. Then, I will try to defend a thesis of extended self with the help of the notion of “narrative self” which has been developed by Daniel Dennett.

II.

When it comes to the extended mind, the extension will always happen from the center of something. If the external processes or states can be understood as part of an extended cognitive process, there should be a subject of these processes and states at the center of such extension. It means that there must be an “I” or the mind of an “I” at the center of the extensions, and external processes should be extensions of the “I” or the mind of the “I.” However, this implies that the cognitive process belonging to me exists outside of my body. The cognitive process ascribed to me or possessed by me is definitely a part of my “I.” When we discuss an ordinary object at the macroscopic level, the object is located where and only where its parts are located. If we apply such a standard for the spatial location of a quotidian object to cognition, the result is that “I” as a self is located wherever the cognitive parts of “I” are located. Doesn’t this amount to the claim that I am located outside my body? If so, I am not confined within my body. I exist outside my body in a form of distributed states.

LYCEUM

When it is understood in such a way, the extended mind not only suggests a new way of interpreting the roles of our body or the world in cognition, but it also seems to demand reinterpreting the notion of the self in a dramatic way. “The extended mind,” simply referring to the extension of mental states or processes, should be distinguished from “the extended self,” postulating the extension of a self as the subject of those states and processes. As stated in the above quotation, however, Clark and Chalmers think that the idea of the extended mind implies the thesis of the extended self. Such an implication seems to be taken for granted.

Eric T. Olson (2011), however, argues that the extended-mind thesis does not directly imply the thesis of the extended self. For example, if it is possible that any cognitive processes or states belonging to me can be realized by something that is not part of me, the extended self will not follow from the extended mind. Olson thinks that Descartes’ substance dualism can be regarded as asserting such a position. If Descartes’ theory is true, “I (my mind)” is clearly distinct from my body. It being admitted that my memory is stored partly in my brain, then, the cognitive states belonging to me can definitely exist outside the border of a non-material substance, which is “I (my mind).” Thus, in order to proceed from the extended mind to the extended self, Olson insists that there should be the additional premise that “A being’s mental states can never extend beyond its own boundaries,” and he calls this “mental-state internalism” (484). According to this mental-state internalism, the mental states of a being must be located entirely within it.

Olson argues that mental-state internalism should not be taken for granted; it needs to be justified. Contrary to Olson, I think that mental-state internalism is a view protected by the principle of presumption. In other words, I think it is those who reject this principle who should take the burden of proof, not those who accept it. As long as no one suggests a critical ground for rejecting the it, we should assume that mental-state internalism is true. Olson refers to the scholars like Lynne Rudder Baker and himself as the philosophers who resist mental-state internalism. I believe that they fail to give proper grounds for denying mental-

Extended Mind and the Extension of a Self

state internalism. Even though Olson and Baker offer different rejections, they both adopt a sort of organism-centrism when they characterize a self or an agent.

Let's first take a closer look at Olson's claim. Olson's strategy to show the implausibility of the thesis of the extended self is to infer a seemingly unacceptable conclusion from the thesis of the extended self. We may reconstruct his basic argument as follows. Olson first tries to prove that Otto is not a biological organism (p. 486):

If Otto is an extended self, then he will be expanded or reduced depending on whether or not he uses his notebook. (Otto's notebook is not always a part of Otto. If he stopped using the notebook, then it would not be a part of Otto anymore.)

The biological organism O, which is in fact Otto's body, will not be expanded or reduced based on whether it uses the notebook. (Even if the notebook is a part of Otto, the notebook cannot be a part of O but would belong to the environment of O.)

Therefore, Otto is not the biological organism O. (Otto \neq biological organism O.)

Moreover, since Otto has the modal property *having a notebook as a part at some time*, while the organism O could not have such a property, Olson draws the stronger conclusion that Otto is essentially not an organism, nor is he contingently or even temporarily an organism. Generalizing this, he draws the conclusion that anyone who possibly has external mental states, even at times when all of his or her mental states are internal, is not an organism. This is because an organism cannot have such a modal property. In a nutshell, every being capable of having an extended self is not an organism.

LYCEUM

This may seem a very radical contention, namely, to argue that any mental being who can have an extended self as Otto can is not an organism. However, if we reflect further, then we see that, contrary to Olson's intention, this is the ultimate implication of the extended-mind thesis and the contention is not as extreme as it initially looks. The sentence denying the identity between Otto and the organism O implies that different conditions of persistence or of identity should be used in individualizing them; it does not imply, however, that Otto and the organism O are separate beings. For example, Otto can be individualized in terms of having a first-person perspective or of having the capacity or responsibility for his actions; by contrast, the organism O will be individualized through its biological composition or functions. This is compatible with the fact that Otto is extensionally identical with the organism O at some point before he uses his notebook. What Olson intends to say via the above sentence (3) can be formalized as the following: $\sim\forall F(F_{\text{Otto}} \leftrightarrow F_O)$. This is logically equivalent to the formula $\exists F\sim(F_{\text{Otto}} \leftrightarrow F_O)$, and it claims that there is some property that Otto has but the organism O does not. If we understand sentence (3) in this way, it cannot be any threat to the thesis of the extended self.

However, Olson seems to understand this sentence as follows: $\forall F\sim(F_{\text{Otto}} \leftrightarrow F_O)$. Based on this, he develops additional criticism of the thesis of the extended self. One of the critiques is, as he calls it, the problem of too many thinkers. We do not deny that a normal, matured human organism, which is identical with our body, is capable of thinking. According to Olson, however, if we admit that "I" as an extended self and my physical body (an organism) are different and that the organism can be capable of thinking, then there will be two beings who can think and write this paper. As it were, there are two thinkers where there should be only a single thinker: one is "I" as a person, not an organism; the other is an organism distinguished from the person. This conclusion is unwarranted because, as noted earlier, Olson misunderstands sentence (3), which is the negation of identity. As noted earlier, what sentence (3) says is no more or less than that the individuation or identity condition of a self (or person) as the subject of thinking is different from the individuation or

Extended Mind and the Extension of a Self

identity of an organism. The relationship between Otto and the organism O, as it were, is the one between a whole and its parts, the relation of *constitution*. A lump of bronze constitutes a statue made of bronze, but the relation of constitution is not the same as the relation of identity. A bronze statue has a certain shape at a certain point. At that point, the lump of bronze has the same shape as the statue. At this moment, are there two different objects which have the exact same shape? Olson would seem to think so. I disagree. On my view, there are simply two different criteria to individuate an object, from which, the existence of two numerically different objects does not follow.

III.

In comparison with Olson's view, Baker's is much more moderate. In a certain respect, the conflict between the extended-self thesis and Baker's position is more a linguistic dispute than a substantial difference. First, let us examine what Baker's argument against the extended self is about. The key point of Baker's argument centers on the conception of a person (a self). Baker (2009) states that a person or a self is an agent with an integrated body or an enduring subject of experiences. This person is a material being constituted by a body (organism). A person and a body (organism) have different persistence conditions, such that they are not in the relation of identity. A person can persist as itself so long as it maintains a first-person perspective, which amounts to having the ability for thinking reflectively about itself. By contrast, an organism has a third-person-persistence condition in that it persists insofar as it maintains certain biological functions. Therefore, a person is not an intrinsically biological being; an organism, by contrast, is biological in itself. If an organism has a first-person perspective, then it can become a person. The combination "organism with a first-person perspective" is a contingent relation emerging from a long history of evolution. The relation between a person and a first-person perspective, by contrast, is essential.

LYCEUM

According to Baker, the relation between a first-person-perspective person and an organism is that of constitution, and their combination is contingent, so a person can have non-organic bionic components as parts that constitute it at the sub-personal level. Non-biological components can also play a critical role in the cognitive processes undergone on the sub-personal level. Considering the tendency for the line between human organisms and machines to blur as technology develops, Baker does not exclude the possibility that a person may one day in the future be constituted by an integrated body consisting of an organism and various prosthetic devices or by a nonorganic body consisting purely of artificial organs and prostheses. This means that non-organic, artificial devices may be material vehicles for the mental states of a person. Furthermore, Baker agrees with vehicle externalism, arguing that a person can be extended cognitively. For example, Otto can have mental states or processes implemented by non-biological vehicles like a notebook. Therefore, she entirely agrees with the extended mind thesis.

What Baker disagrees with is the move from the extended mind to the extended self. She concurs that there is an integrated cognitive system including Otto and his notebook as parts. But even though she thinks this, she holds that Otto cannot become an extended self, expanded to include his notebook as part of himself. According to Baker, Otto as a self (or person) is simply not a cognitive system. Otto as a person is a concrete particular with a first-person perspective, constituted by his body. The coupling of Otto and his notebook or his body and his notebook, however, cannot constitute anything; no concrete particular is denoted by "Otto and his notebook." In that respect, Otto has ontological priority over the integrated system (Otto-cum-notebook).

Baker thinks that accepting the extended mind does not commit one to the claim that a self as a persistent person is also extended; the only cognitive subject is just Otto constituted by his body. In other words, Baker thinks that a person is not only constituted by a body but also spatially congruent with the constituting body; a person does not extend beyond the body. Of course, organic bodies can gradually turn into artificial ones through the integration of organic

Extended Mind and the Extension of a Self

parts and machines. The material basis constituting a person can be altered; however, a person will continue being identical only as far as a first-person perspective is sustained. She also believes that the change of such a material basis would not extend the “I” as a persistent person (agent, self) into the environment. As a person, I am an agent and the subject of experience, not simply an integrated cognitive system or one of the constitutive components of it. In addition, an extended self beyond its body does not exist. In a nutshell, what Baker contends is that the skin of an organism is not a boundary for its cognitive vehicles, but it is one for a person constituted by a body.

I characterized both Baker’s and Olson’s views as a sort of organism centrism. To be more precise, Baker’s view is more a body centrism about the location of a self. That is, the location of a self is confined to its body, which constitutes a person.¹ What is Baker’s argument to show that the location of a person is confined to the body? It is not clearly stated in her paper, but she seems to think that only a body is a boundary in terms of which a first-person perspective can be talked about meaningfully. According to her, a person exists as long as a first-person perspective is maintained. The minimum material basis implementing a first-person perspective is a person’s body, whether organic or artificial. When the body that constitutes me no longer exists, “I” cannot exist anymore, either. For this reason, she seems to think that a body is the physical boundary for its self.

Baker’s contention that the boundary for a self as an agent is demarcated in terms of its body, the physical basis of the self, is in good accordance with our ordinary intuition for our current social practices, such as the subject of legal liability. However, as Kurzweil holds, if mind uploading becomes possible—that is, if a human’s self can exist as a sort of disembodied informational structure—

¹ According to Baker, bionic-devices permanently integrated to my body may constitute parts of me. Artificial bionic-devices can be parts of a person’s body only if the following conditions are met: (a) it is causally integrated with the other parts that maintain the functioning of the body, and (b) it is permanently in place either inside the skin or attached to the skin on the outside (pp. 654-55).

LYCEUM

what would we think of it? A self as an informational being may exist not as permanently combined with a particular body but as changing its physical body freely as occasion demands. Or several different selves may share a common physical foundation, which is not divided separately. If such is the case—that is, if our ways of existence are fundamentally altered—what would be the consequence? Would such a situation mean the end of selves (persons)?

In reality, even though they are not so extreme, we already encounter various phenomena that are at odds with the notion of selves based on organic bodies. For example, in hyperspace, such as the Internet at large or Social Network Services (SNS) more narrowly, we may emotionally bond with and get social recognition from others, not via physical contacts but through online activities only. Furthermore, in hyperspace, it is possible to have multiple personalities, each an “I” with very different sexual, political, and/or social identities from the ones we have in the offline world. Thanks to these electronic communities, there are new types of communication, contacts, and presence; and other selves have emerged that are fitted to these highly different conditions, limitations, and possibilities. In addition, via texting and online-chatting, we are constantly “contacting” people being or living far away. At the moment, we exist not only where we are physically located, but we are also “present” in some sense where we are connected remotely through an electronic network. For the time being, this “tele-presence” is implemented only by using some restricted, limited sensory modalities, such as in video conferences, voice chatting, or texting; in the future, however, tele-presence in the form of an alternative body, like a physical avatar, may very well become prevalent.

What are the implications of such phenomena on the notions of self or person? Baker focuses on the minimum physical foundation for a person to be sustained with a first-person perspective. When Clark and Chalmers talk about the extended self, I think that they would not deny the role of a physical body as the minimum material foundation for the *persistence* of a self with a first-person perspective. Their focus, however, is on what the physical foundations for the *personal characteristics* of a self are, which are identified in terms of what we

Extended Mind and the Extension of a Self

do, what we are capable of, who and what we think we are, what personality we have, and what we express about ourselves. The main point of the extended-self thesis is that my nature or various characteristics that I have as a person are not determined or limited by the minimum physical foundations for sustaining a first-person self; the vehicles of these traits are extended beyond the boundaries of my physical body. Thus, the significance of the extended-self thesis lies in revealing the deficiency of Baker's notion of the self, which is confined to a physical body, in its power to explain or understand the new phenomena emerging from the development of communication and scientific technology.

IV.

In that respect, the thesis of the extended self does not just extend boundaries for the self but tends toward overthrowing notions of self and person as we have traditionally conceived them. In an orthodox view, a self is equated with the subject of unitary conscious experience, and this self is set against the external world. This reflects the Cartesian conception of self as a *res cogitans*, which controls my body and my will and makes plans and decisions. This is also the place in which my beliefs, memories, values, and personality reside. Most contemporary philosophers do not accept such a conception of self or non-substantial soul as Descartes did. Nevertheless, many philosophers still believe that there must be some physical center which takes over the role of Descartes' posited self. They still ascribe conscious thoughts and mental activities to this thing-like self, and maintain a substantial notion of self, according to which a self has shape and occupies location. Since Baker conceives of a self as being intimately combined with its body, she also seems to belong to this tradition in which a self is closely affiliated with some physical center.

The extended-self thesis challenges this conception of self, suggesting that the notion of self belongs to a wholly different category from that of substance. According to Clark, the notion of "self" indicates a flexible and open system including non-biological, external devices as its parts. He holds that the

LYCEUM

characteristics making us essentially human as we are now are those that collaborate with external, non-biological devices and resources and that are willing to adjust our activities to them in order to find a better solution for the problems of survival and reproduction. In that respect, our selves cannot be something internal confined to our bodies or brains. Human selves are a system of reasoning, inference, and action situated across the boundaries between biological brains and non-biological circuits; they can by no means be restricted within boundaries like biological skin or skull. Even though the symbiosis among humans, devices (technology), and the environment is constructed and defined by human beings, it also, at the same time, constrains or redefines our potentials at the fundamental level. As a result, human beings have become a sort of hybrid being that reveals its full identity only as a combination of biological brains, bodies, intelligent devices, and technology.

However, the self we discuss here is no longer a substantial thing. The self is now a cluster of various processes and activities going along under the name of “self”; the notion of the enduring subject of experience is substituted by the notion of “narrative self,” which is constructed of various interactions happening at the cognitive mechanism. In “The Origins of Selves” (1989), Daniel Dennett finds the uniqueness of human beings distinguished from other beings such as animals in their ability to represent themselves through language. In the case of other animals, the boundaries between self and external world are organized by skin (as in tiger), shell (as in lobster), or colony (as in ants). Meanwhile, humans create boundaries with stories about themselves. Spiders protect, control, and define themselves by weaving webs, and beavers by building dams. Unlike them, however, we humans organize the boundaries between ourselves and the external world by making stories. As it were, we are protecting, controlling, and defining ourselves by narrating stories continuously to ourselves or others about whom or what we are:

And just as spiders don't have to think, consciously and deliberately, about how to spin their webs, and just as

Extended Mind and the Extension of a Self

beavers, unlike professional human engineers, do not consciously and deliberately plan the structures they build, we (unlike professional human story-tellers) do not consciously and deliberately figure out what narratives to tell and how to tell them. Our tales are spun, but for the most part we don't spin them; they spin us. Our human consciousness, and our narrative selfhood, is their product, not their source. These strings or streams of narrative issue forth *as if* from a single source—not just in the obvious physical sense of flowing from just one mouth, or one pencil or pen, but in a more subtle sense: their effect on any audience or reader is to encourage them to (try to) posit a unified agent whose words they are, about whom they are: in short, to posit what I call a *center of narrative gravity*. (Dennett, 1989, p. 169)

For Dennett, in short, a self is the center of narrative gravity as a sort of theoretical fiction constructed out of a variety of stories that constitute the chronicle of a certain living body, a self that provides unity and coherence to the stories. What the extended-self thesis is trying to say can be more properly understood via Dennett's notion of the narrative self.

As stated above, Chalmers asserts that his iPhone is a part of himself. However, if an iPhone is a proper part of him, constituting his mental processes, then the processes producing and controlling his behavior would consist of states scattered across the boundaries from his brain or skin reaching to the external object called "iPhone." If such is this case, where is his self located? In what state or in what shape is it? Does it exist as distributed as his mental processes do? Does his self literally exist as a physical combination of his body (or brain) and iPhone? If so, when he leaves iPhone in the bedroom, does he exist in distributed states both in the living room where his body is located and in the bedroom where his iPhone is located? We may feel a bit perplexed by these

LYCEUM

questions even before we try to answer them. This is not because the questions are difficult, but because we cannot be so sure whether the questions are the right ones to ask.

The proponents of the thesis of the extended self consider such questions as wrong queries arising from a category mistake that places the notion of self in the wrong conceptual category. That is, these questions arise from a Cartesian legacy that regards the self as a sort of thing having a location and form. The notion of narrative self, avoiding these puzzling questions, provides the theoretical basis on which self can be characterized through the symbiosis of humans, tools, and the environment. According to this notion of narrative self, self is not a material thing to which our various mental characteristics are ascribed, or a place in which our mental activities are occurring. The concept of the extended self as a narrative self does not identify the self with a physical center of unitary consciousness that is composed by all of our thoughts, but urges us to liken it to the center of a narrative binding diverse stories about ourselves, told by ourselves and others, into a coherent story. The extended self is a sort of thread that runs through the tasks we carry out, the goals or plans we try to accomplish, and the responsibilities we take on; this cannot be a center of cognitive processes that occupies a space. Clark says:

I think of myself not just as a physical presence but as a certain set of ongoing goals, projects, and commitments: to write a new paper, to be a good husband, to better understand the nature of persons, and so on. These goals and projects are not static, nor are they arbitrarily changeable. I recognize myself, over my lifetime, in part by keeping track of this flow of projects and commitments. Others, likewise, will often recognize me as a unique individual, not (or not only) by recognizing my physical shape and form but by recognizing some distinctive nexus of projects and activities. (2003, p. 132)

Extended Mind and the Extension of a Self

Such a narrative self is the self confirmed in our own and others' mutual story. That is, this is a self made up of what we and others think about our goals, projects, abilities, and potential. If such a narrative self is one way of defining our own selves, the narrative self of the digital age will be a collection of stories made by a sort of bio-techno hybrid consisting of our brains and the intelligent environment we depend upon.

Through the advent of an intelligent environment, there would no doubt be a great change in our goals and projects, our abilities, and our potential. For example, even in our present day, when we want to solve a problem, we rely on various external tools. As I am writing this paper, I am using various books and articles, memos, electronic files, software, and search engines. The engine that actually solves a problem is therefore not just my brain but the matrix that includes various instruments offering such technical assistance. According to the notion of narrative self, these instrumental matrices are important components of my stories or activities, such that these become prominent parts of me to define who and what I am and what I am capable of.

Against this claim, some may offer the critique that a narrative self and the extended self are merely exaggerated rhetorical expressions used by technophiles, for our brain ultimately exerts the authority to command, and devices like smartphones remain passive, awaiting our direction. Dennett (1984) has argued, "I am the sum total of the parts I control directly," contending that control is the ultimate criterion (p. 82). The brain appears to control my behavior and choices in asymmetrical ways that cannot be done by pens, paper, computers, or smartphones. If control is the ultimate standard by which to define a self, shouldn't we say that our selves are biological brains or bodies in the traditional way instead of the hybrid of life and technology?

This point sounds relatively reasonable when the roles of devices clearly remain passive, as in using paper and pencils to calculate. However, as indicated by the advent of current state-of-art digital technology, if the tools or the environment get more intelligent, our brains will, more and more, delegate to the intelligent environment the burden of many cognitive processes related to

LYCEUM

controlling our behavior. For example, think about an e-book store like *Amazon*. Since hundreds of new books are published and uploaded to its database every day, it is almost impossible to look into the entire inventory of books stored in *Amazon* no matter how much time we invest in checking it. However, *Amazon* applies an algorithm that analyzes patterns of browsing its website or of my previous purchasing behavior, compares my patterns with other consumers who have patterns similar to mine, and then suggests books that I may be interested in. In many cases, instead of searching every book in the prodigious databases, I often purchase the books that look interesting from the list that *Amazon* suggests. In this case, where does the power of controlling my choices really come from? In such a case, since my biological brain is harmoniously working with the intelligent environment in a seamless way, it is not easy to distinguish where my intervention to choose and command begins. The more these technologies are developed, the harder it will be to discern the boundaries or determine the factors that produce my behavior.

Clark criticizes invoking the notion of control in an attempt to reinstate the biological brain in its post as a self. He claims that such an attempt would understand the notion of control in a very restricted and strong sense, such that it would shrink the self into something extremely meager, regarding the intelligent environment as a mere external environment, so that:

It would be as someone tried to argue that the “real me” excludes all those nonconscious neural activities on which I so constantly depend relegating all this to a mere smart inner environment. . . . The vision of the mind and self that remains following this exercise in cognitive amputation is thin indeed! . . . The intelligent system that now confronts the wider world is biological-you-plus-the-software-agents. These external bundles of code are contributing as do the various nonconscious cognitive mechanism active in your own brain. They are constantly

Extended Mind and the Extension of a Self

at work, contributing to your emerging psychological profile. You finally count as “using” the software agents only in the same attenuated and ultimately paradoxical way, for example, that you count as “using” your posterior parietal cortex. (pp. 30-31)

However, because the final decision seems to come from our *conscious* judgments, isn't it possible to say that the brain is still the center of self? In *Elbow Room*, Dennett provides a way to respond to such a rebuttal by pointing out that many of our cognitive activities take place in the sub-personal processes. He claims that the roles of conscious mind on the personal level are just those of the manager who sets a goal and creates and maintains the conditions in which the component members may contribute to accomplish the goal, but the conscious mind does not exert any privileged control over them. The source of trouble here is the very idea that diverse neurological or non-neurological processes require a privileged user who enjoys a perspective to access and control them. If we limit the boundary of a self within what makes the ultimate decisions, shouldn't a self be identified with frontal lobes? Or if any part of the brain does not have the authority for final decisions, then have the mind and self simply disappeared?

According to Clark, different neural circuits of the brain provide different capacities, and all contribute in different ways to our sense of self, of where we are, of what we can do, and to decision-making and choice. Furthermore, non-biological elements outside our body offer us the possibility of having much more capability than the biological brain alone is provided with; as a result, they provide still further capabilities and contribute in additional ways to our sense of who we are, where we are, what we can do, and how we make decisions and choices. No single process within this complex matrix thinks intrinsically by itself or has an ultimate controlling power, such that any of it can be “the seat of a self.” What makes a self, an individual person, is shifting coalitions consisting of biological circuits and tools. Therefore, Clark says, we are “soft selves” who

LYCEUM

are “continuously open to change and driven to leak through the confines of skin and skull, annexing more and more nonbiological elements as aspects of the machinery of mind itself” (p. 137)

Let me summarize what I have done in this paper. According to mental-state internalism, a being’s mental states can never extend beyond its own boundaries, but according to the extended-mind thesis, our mental states and processes can extend beyond our body into the environment. According to the latter if and when mental states and processes extend beyond our skin, we, as the subjects of such states, also extend beyond our skin. This is the thesis of the extended self. In order to show the plausibility of this thesis, I examined the objections of Eric T. Olson and Lynne Rudder Baker and tried to elucidate the notion of extended-self in terms of “narrative self.” The extended-self thesis, thus understood, challenges the traditional conception of self, which views a self as a sort of substantial thing that possesses our various mental activities or characteristics. According to the conception of narrative self, a self is understood as the center of narrative gravity for the stories told both to ourselves and to others, and told both by ourselves and others, consisting of a certain set of ongoing goals, projects, and commitments.

Why is it important to have such a notion of extended self? I believe that the scientific technologies of today, such as digital technology and bioengineering, are technologies that redefine and redesign ‘human being’ itself. As we converge more with digital tools and prosthetic devices, the boundary between the self and tools as traditionally conceived cannot remain intact. I predict this will inevitably force a radical change in the way we understand the concept of agency and selfhood. Thus, the convergence of human beings and state of the art technology currently in progress is not simply about the development of technology; it concerns a fundamental question about the human being. It will have various implications for our current cultural, moral, and political practices. For example, when we want to hold someone morally or politically responsible for a consequence, who should we turn to? Would the subject of responsibility be the ‘individual’ who is confined to the biological

Extended Mind and the Extension of a Self

body? Or is it the hybrid fused with the environment? Is it a coherent idea at all to hold the hybrid responsible? Or is the awkwardness we feel with such an idea just a legacy of the traditional conception of self? Answering these questions would require us to do more develop and elaborate the notion of extended self.²

References

- Baker, L. R. (2009). Persons and the extended-mind thesis. *Zygon: Journal of Religion and Science*, 44(3), 642-58.
- Chalmers, D. (2008). Foreword. To A. Clark (2008). *Supersizing the mind* (pp. ix-xvi). Oxford: Oxford University Press.
- Chalmers, D. & Clark, A. (1998). The extended mind. *Analysis*, 58, 7-19.
- Clark, A. (2003). *Natural-born cyborgs*. Oxford: Oxford University Press.
- Dennett, D. (1984). *Elbow room*. Cambridge, MA: MIT Press.
- Dennett, D. (1989). The origins of selves. *Cogito*, 3, 163-73.
- Olson, E. T. (2011). The extended self. *Mind & Machine*, 21, 481-95.

² This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2007-361-AL0015).

Two Sorts of Self-Creation: On Galen Strawson’s “Basic Argument”

Griffin Klemick

In a handful of works over the past few decades, Galen Strawson has formulated and defended an argument he calls the “Basic Argument” against moral responsibility.¹ This argument leverages the seemingly demonstrable impossibility of control over one’s character against any account of responsibility robust enough to ground our moral practices. Strawson has given the Argument in myriad formulations, but one has become particularly prominent in the literature. I give this version below; in my view, it omits nothing central to Strawson’s argument.²

The Basic Argument

- (1) In any given situation, you do what you do because of the way you are.
- (2) So in order to be ultimately responsible for what you do, you must be ultimately responsible at least for particular mental aspects of the way you are.³
- (3) But you cannot be ultimately responsible for the way you are.

¹ I will always give *Basic Argument* in capital letters. Where *Argument* is given with a capital letter, it refers to the Basic Argument as given below.

² The first four premises of the argument are based on Strawson (2002, version 1); the regress component is based on Strawson (1994, sec. I).

³ The qualifier “at least for particular mental aspects of” is added to forestall the idea that one must be responsible for *all aspects* of oneself—for one’s hair color, ancestry, intellectual aptitude, and so on. The mental aspects in question are characterized in premise 6. I will take this qualifier for granted in the remainder of the argument.

***Two Sorts of Self-Creation:
On Galen Strawson's "Basic Argument"***

- (4) Therefore, you cannot be ultimately responsible for what you do.

Why can't you be ultimately responsible for the way that you are? Because:

- (5) To be ultimately responsible for the way you are, you must have consciously and explicitly chosen to be the way that you now are. Call this choice "C."
- (6) But for you to have *chosen* to be the way you now are, you must have existed prior to your being that way, and at that prior time you must already have held to certain principles by which you might choose C—preferences, values, ideals, and the like. Call these principles collectively "P." (In what follows, I will call these principles one's *volitions* and their collection one's *volitional makeup*.)⁴
- (7) But then to be ultimately responsible for C, you must be ultimately responsible for having held P.
- (8) But to be ultimately responsible for having held P, you must have consciously and explicitly chosen to hold P. Call *this* choice "C*."
- (9) But to have chosen C*, prior to C* you must have existed and held to certain principles by which you

⁴ As I will use the term, one's *volitions* include one's first-order desires as well as one's higher-order desires and deep concerns, values, and ideals. Perhaps this use of the term is non-standard, but another lump term was not forthcoming.

LYCEUM

might choose C*. Call *these* principles collectively “P*.”

- (10) This cycle—ultimate responsibility for mental aspects of yourself requiring choosing those aspects, and ultimate responsibility for those choices requiring an appeal to still prior mental aspects—is a backward infinite regress.
- (11) So to be ultimately responsible for the way that you are, you must have actually made an infinite series of choices.
- (12) But it is impossible for a finite being to have actually made an infinite series of choices.
- (13) Therefore, it is impossible to be ultimately responsible for the way that you are.

For some time, I found myself in the position in which Berkeley left Hume: the Argument produced no conviction in me, but admitted of no decisive response. I now think this stemmed from my misconception of what the first step toward a response would have to be. Strawson’s argument ultimately derives its support from an appeal to intuition: the notions of control over self and moral responsibility to which the Argument makes reference are “central to ordinary thought about moral responsibility and justice” (Strawson, 1994, p. 222). For this reason, it seemed to me that compatibilists should argue that our intuitions do not support the requirement of Strawson’s “ultimate responsibility” for moral responsibility, for all the reasons given by the standard arguments for

Two Sorts of Self-Creation: On Galen Strawson's "Basic Argument"

compatibilism. And libertarians should argue that ultimate responsibility is possible,⁵ for all the reasons given by the standard arguments for libertarianism.

I see matters differently now. To be sure, the fact that libertarians and compatibilists can so easily leverage their standard arguments against Strawson's premises does call into question the Argument's dialectical value. But framing the issue in this way allows a potential deeper point of agreement between libertarians and compatibilists to be drowned out by the familiar "dull thud of conflicting intuitions." In this paper, therefore, I wish to consider a more fundamental answer to the Basic Argument, one arrived at through a less direct route. I propose to examine in some detail two accounts of self-creation—one expressed by the Basic Argument; the other, a more modest proposal, detailed by Harry Frankfurt and Charles Taylor—and to consider which sort of self-creation places us in a better position to attain the values we seek through our agency. Does each form of self-creation allow us to identify with our selves, or does it leave us alienated from our volitions and our actions? In answering this question, I will develop a challenge to Strawson's claim about our ordinary conception of responsibility based on premises both libertarians and compatibilists can accept.

First, I will give an account of what it is to be *causa sui* and a preliminary description of the goods the *causa sui* agent can and cannot achieve. Next, I will sketch an alternate account of our ability to shape ourselves, arguing that our particular actions and our attempts to articulate our fundamental values imbue us with significant control over our wills. Finally, I will argue that the latter form of self-shaping better positions us to attain the goods desired in agency, precisely because it situates self-creation within human contexts and communities.

⁵ Or perhaps they should argue that intuition requires only a less demanding type of incompatibilist responsibility, which type we actually possess.

LYCEUM

Strawsonian Self-Creation

For Strawson (1994), if we are to have a meaningful measure of control over our selves, we must create them—we must be *causa sui* with respect to them (see p. 212). Obviously, this formulation is unhelpful by itself: we need a detailed account of what it is for human agents to be self-creators in the relevant sense.⁶ We may look to the regress component of the Basic Argument for a concrete illustration: what is required is that, for every aspect of an agent's volitional makeup, she has consciously, explicitly and reflectively chosen that aspect, and at least one such choice is not motivated by any further aspect.⁷ It seems appropriate to impose further affective conditions: let us say that the agent must care about which kind of will she comes to have, and she must recognize her responsibility for her choice (rather than, e.g., ducking it through Sartrean bad faith). If these conditions are met, we have a self that carefully and deliberately builds itself up from nothing. We might think that the “finished product” is a self for which the agent can rightly feel ultimately responsible.

Note that these conditions actually go beyond the requirements described in the Basic Argument's regress scenario. This is necessary: while the Argument

⁶ Strawson's parody of overly general accounts is apt: to be *causa sui* is “really quite simple . . . in the case of stones, for example, it's just for the stone to be—truly—the origin or creator of itself” (2004, p. 362). Obviously, this doesn't help us understand why we should identify with the self we create.

⁷ A few points on this rough summary. First, concerning the requirement of conscious, explicit reflection, Strawson tends to restrict the scope of the Basic Argument to “fully intentional and consciously deliberated actions” (2002, p. 442), but for expository rather than theoretical reasons. So it might seem that this requirement goes too far. But Strawson views uncaused actions as random occurrences for which no agent can be responsible (see his 2000), without apparent regard for whether the absence of causality occurs at the psychological level or the physical one. Second, since the Basic Argument is challenging responsibility for actions, it requires that an agent be responsible only for those aspects of the agent's will that issue in or contribute to subsequent actions. But since the issue here is with identification with the self, it seems reasonable to extend the requirement to all aspects of the agent's will. (In any event, how can a volition be a part of an agent's self in any deep or important way if it has no effect at all on her behavior?)

Two Sorts of Self-Creation: On Galen Strawson's "Basic Argument"

takes its cues from the problem of being *causa sui* in time, the property of being *causa sui* is purely *structural*; it has no essential *temporal* components, and so the condition that all aspects of our will be chosen is not sufficient for our being *causa sui*.⁸ For an agent who, in an infinite series of choices, chooses every aspect of herself may yet fail to be *causa sui*. This could be true, for instance, if she made each choice in this infinite series of choices with reference to "still prior" decision-making criteria; on this account, the structural question of whether the agent controls her will or is controlled by it is unclear.⁹ On the other hand, it seems that an agent who exists entirely outside temporal boundaries could be *causa sui*, if the right structural relationship obtained between her volitional makeup and her choice—if, in one timeless act, she eternally chooses all her values and desires without reference to further, un-chosen principles of choice. (Strawson seems to recognize this when he softens his claim that "nothing can be ultimately *causa sui* in any respect at all" to the claim (1994, p. 224) that "even if God can be, we can't be.")¹⁰ To be *causa sui* in the Strawsonian sense, then, all an agent's choices must structurally stem ultimately from only the choice of the agent's "bare self."

⁸ This does not, of course, mean that Strawson is wrong to draw out the impossible necessary conditions of a *temporally bound human's being causa sui*. It is only to note that these conditions are particular to humans, and that the property is not itself temporal in nature.

⁹ Of course, it is unclear what we should say about this agent's control over her will, given the introduction of the impossible actual infinite sequence. But we cannot for this reason disallow the example, since the whole point of the Basic Argument is that key human goods depend on our possession of a kind of control over ourselves that is achievable only by meeting impossible necessary conditions. It is only fair, then, to ask whether the impossible necessary conditions would actually help us achieve any key human goods. And this seems in part to hinge on the unanswerable question whether it is more important, in cases like the one above, that *the agent chooses all her reasons* or that *each of her choices is controlled by a prior reason*. I contend that in such examples, there is no reason to assume that the first fact is more important.

¹⁰ And even if God can't be *causa sui*, I doubt that this is due to God's transcendence of temporality.

LYCEUM

What sort of self can take its place at the top of this structure, in which all aspects of its will depend ultimately on reason-less choice? Only a bare subject of experience, which Thomas Nagel (1986) describes as a “pure, featureless mental receptacle” (p. 33), fits the bill.¹¹ This self is to the mental aspect of a human being what prime matter is to a human body. Like prime matter, such a self cannot actually exist, and some compatibilists have argued that this should gainsay the claim that it figures crucially in meaningful agency: for instance, Daniel Dennett (1984a) contends that “if the ideal of freedom we hold out for is simply self-contradictory, we should hardly feel bereft when we learn we cannot have it. There’s no sense wringing our hands because we can’t . . . create ourselves *ex nihilo*” (p. 172).

In one sense, this is quite right: it is *useless* to regret our inability to be *causa sui*,¹² since it’s not our fault we lack this ability and we have no prospects of acquiring it. But perhaps this does not mean that we cannot intelligibly wish that we had it or that we cannot recognize the goods we might attain by it. To borrow an analogy from Strawson (2004), it is only because we can state with precision what it is to be a square circle that we know this is impossible (p. 360); likewise, he contends, we know that being volitionally *causa sui* is impossible because we understand roughly what it is to have this ability and, perhaps, what further goods it would confer on us. In view of this understanding, we can intelligibly wish that we could achieve these impossible goods, and we may feel alienated from ourselves because we lack them.^{13,14} At any rate, I will assume

¹¹ Cf. Nagel’s attempt to identify this subject: “I know what I mean by ‘I.’ I mean *this!*” (1986, p. 33).

¹² Cf. R. E. Hobart (1934, p. 18).

¹³ For a compelling presentation of these points, see Nomy Arpaly (2006, pp. 126-33).

¹⁴ One might object that, while we can understand the goods that our possession of nomologically or otherwise contingently impossible abilities would confer on us, we cannot state with certainty the goods that logically impossible abilities would confer on us. The claim *if I had the ability to be in two places at once, I could go to more parties than I can* seems straightforwardly true, even though its antecedent is impossible. By contrast, it is totally unclear how we should determine the truth or falsity of the claim *if my basketball were shaped like a three-dimensional*

Two Sorts of Self-Creation: On Galen Strawson's "Basic Argument"

that we can determine what goods the ability to be *causa sui* might confer on us and will concern myself solely with the question of what goods these are.

It seems doubtful that the ability to be *causa sui* could confer on us any goods at all. Since the bare subject lacks any volitions by which it could form preferences, it is not clear how it could choose anything at all (or how it could *understand itself* to have this power of bare choice, which might be necessary for it to have a sufficient measure of control over its exercise of this power).¹⁵ Further, even if the bare subject could make choices, there seems no reason to think that it would have the tendency to make *good* choices, for it lacks criteria for distinguishing the good from that which is not good. So the genuinely *causa sui* self cannot dispel our worries about our ability to be moral agents or to stave off alienation.¹⁶

circular square, I could make more impressive bank shots. And the claim if I were causa sui, I could better identify with my will seems more like the latter than the former.

I think this objection is correct, but I would prefer not to make it bear the weight of my whole argument, particularly when the argument in the text potentially yields even more conclusive results by granting Strawson even more and yet confuting his argument.

¹⁵ Hobart (1934) points out that even if an agent could be *causa sui*, we would nevertheless have difficulty *understanding him* to have this ability: "If there were a being who made his 'original character,' and made a fine one, and we proceeded to praise him for it, our language would turn out to be a warm ascription to him of a still earlier character, so that the other would not have been original at all" (p. 18). I am suggesting that it is possible that not only the observers but also the agent would have this difficulty.

¹⁶ One might respond that there remains a chance that the bare subject will choose the good, whereas some human beings are deprived of even a chance by the influence of heredity and environment. But there are two reasons to reject this response. First, unless human beings are fully incapable of apprehension of moral reasons, they have some resources by which they may choose the good, which the bare subject totally lacks; this seems at least as relevant to assessing grounds for alienation as the chance to which the response appeals. Second, it is not clear that any choice for good the bare subject makes could constitute moral agency, since morality makes claims not only on an agent's choices but on her reasons for choosing. Since a bare choice just is a choice not made for reasons, it is not, *a fortiori*, made for moral reasons either. Charles Taylor's critique of Sartrean radical choice seems to apply just as well to Strawsonian self-creation: by demanding that we choose not only our actions but also the values that motivate them, it "dress[es] up as a moral choice

LYCEUM

Lest the discussion grind to a halt so quickly, however, I think a softened, yet still robust form of “*causa sui* agency” might be plausibly thought to provide us with the goods desired in moral agency. While no self that must choose all its reasons will do, there might be some reasons that a self may take as given without thereby diminishing its control. Perhaps some reasons are binding for all rational agents as such; perhaps, along Kantian lines, we might stipulate that these include moral reasons. Even if this is not a tenable metaethical position, it seems to be an assumption underlying our wish to be *causa sui*, at least as we typically express it in everyday life. For what we ordinarily wish is that we could strip away those particular volitions that hold us back from acting in line with normative reasons, including moral ones, and that we could replace these volitions with dispositions to act according to these reasons. We take for granted that moral reasons will continue to be reasons for us even when we have stripped away many of our actual volitions, and this seems plausible only if moral reasons are reasons binding on rational agents as such.

Let us stipulate, then, that the “bare self” has access to reasons binding for rational agents as such, including moral reasons. This self seems capable of creating a will from which it would not be alienated, since it has an ability to choose that is unencumbered by a particular context and has the required knowledge to comport itself with moral norms. Full self-creation may not be desirable, then, but this slightly restricted self-creation still might be. Two questions remain, then: whether a still more modest form of self-creation is possible for human beings, and whether this still more modest contender is equally desirable—or more desirable—than the slightly restricted version. I argue that both questions should be answered in the affirmative; I treat the former question in the next two sections and the latter in the penultimate section.

what is really a *de facto* preference” (1976, p. 121). If this is right, then totally unrestricted choice deprives one of the capacity for moral agency far more than any typical constraints on agency faced by human agents.

***Two Sorts of Self-Creation:
On Galen Strawson's "Basic Argument"***

The Relationship Between Actions and Volitions

In this section and the next one, I will focus on the simple question of how we might have significant causal influence and control over our volitions if we cannot be *causa sui*. One plausible answer is that we can attempt to articulate our deepest volitions, identify with some of them rather than others, and take steps to comport our actions with the chosen volitions. In the next section I will describe this process in more detail.

But first I want to give this simple question a simpler answer, and so I invite the reader to perform a simple thought-experiment. Simply think of a few of your significant volitions: your interests, concerns, loves and values. Now consider how many of these volitions became so important to you without your actions in any way contributing to the process. Speaking for myself, I find the answer is *none of them*. Of course, I did not typically make the choices that deepened the importance of these volitions to me with full knowledge of how my choices would affect me. Nor, in making any such choice, was I free from the influence of other desires or concerns—still less from environmental influences. Nevertheless, in each case I willingly (and more-or-less reflectively) chose to act in ways that caused my will to take its current shape, and I am convinced this is significant.

Perhaps this answer is not only simple but also shallow. After all, one convinced by the Basic Argument might make the familiar retort: Even if these choices affected my volitional makeup, weren't they in turn caused by other volitions, and doesn't this chain infinitely regress? Perhaps, but I argue that it is equally true that the volitions that influenced my choices did not take shape until I acted in ways that furthered or gave rise to them—the chain of dependence of volitions on actions also infinitely regresses. It may be true that simplistic forms of libertarianism may disproportionately emphasize our ability to choose, considered in abstraction from the influence of volition. But it is equally true that simplistic forms of compatibilism or hard determinism may disproportionately

LYCEUM

emphasize the dependence of actions on one's volitions. Of course, each of my actions stems from some subjective motivation. But there is no reason to think that every such subjective motivation will answer to labels like *my deepest volition* or *my strongest desire*. After all, as Carl Ginet (1989) points out in response to Mill's contention that one's "chosen action will [always] be the one that satisfies whichever of the conflicting motives is stronger than all the others":

One can secure confidence in Mill's law only by making it true by definition: "the strongest motive" *means* the motive that prevails. If the term is defined by some logically independent criterion, so that the proposed law will be a non-trivial proposition, then it is an open question whether the facts would give us reason for confidence in it. (pp. 26-7, Ginet's italics)¹⁷

If one operates with a definition of "strongest desire" such that all subjective motivations that lead to action satisfy the term, this simply evinces the triviality of the established claim.

I suggest, then, that a different model of the relationship between actions, volitions and the self is required: a model that keeps to the middle ground between these two disproportionate emphases. We should reject models that identify the self with the subjective power of decision and locate it above volition, such that its decisions can be made apart from—and can always subsequently alter—the influence of its volitions. We can never choose apart from the influence of volition in general; nor should we underestimate the difficulty of acting to diminish the hold of particular strong volitions over our powers of decision. But equally worthy of rejection are those models that constitute the self from a set of particular, static, mutually consistent volitions and traits, identifying free action with the mechanistic function of these central

¹⁷ The wording of the contention termed "Mill's law" is Ginet's own.

Two Sorts of Self-Creation: On Galen Strawson's "Basic Argument"

volitions.¹⁸ This type of model unduly minimizes the importance of an agent's actions in shaping her volitional makeup; it also underestimates the degree to which the volitions central to her identity may be mutually inconsistent. Interestingly, Galen Strawson blends these two models: for Strawson, the self that is *causa sui* can be nothing but the bare subject, but its actions are a function of its volitional makeup. So, the self must create itself "from the top down": to be responsible for its actions, it must choose (without motives) all the intermediary volitions that will determine them.

I suggest instead that an agent's actions are motivated by given oft-mutually inconsistent volitions, each of which in turn increases or decreases in importance to her based on her actions and experiences. Crucially, neither the will's influence on action nor action's influence on the will is more fundamental than the other. The agent's self does not stand behind her volitions; rather, the self is constituted by the interplay between volitions and actions (as these reflect the influences of her experiences and the communities to which she belongs).

In practicing self-creation through our actions and particular choices, we do not directly create our volitions; they are, rather, self-creation's raw material and indirect consequences. But even if volitions exist in us unbidden, our actions make them a part of us in a more significant way. Harry Frankfurt (1987) grasps this point acutely:

When the decision is made without reservation, the commitment it entails is decisive. Then the person no longer holds himself apart from the desire to which he has committed himself. It is no longer unsettled or uncertain whether the object of that desire—that is, what he wants—is what he really wants: the decision determines what the

¹⁸ Hume seems to have adopted such a model of free agency when he argued that free actions are just those that express the agent's deep or settled character traits. Cf. John Martin Fischer (2005, p. 157).

LYCEUM

person really wants by making the desire on which he decides fully his own. To this extent the person, in making a decision by which he identifies with a desire, *constitutes himself*. (p. 170, Frankfurt's italics)

For Frankfurt (ibid.), it is not the *existence* of the person's desires that is important for moral agency; it is whether the person *takes responsibility* for these desires, and thus makes them a part of his self (see pp. 171-2). And he can do this through his particular actions just as much as by choosing his volitions from the top down.

John Martin Fischer (1999) finds the middle ground between the two disproportionate emphases I've described by contending that the value of free agency consists in the agent's expression of "the meaning of the sentence of the book of his life. And this meaning is fixed in part . . . by the overall narrative structure of the life" (p. 116). In morally responsible agency, we further the story of the self; we express further elements of the interplay between our actions and our choices. In some cases, the new chapter will show how our values, already relatively stable and fixed, worked themselves out in action. In other cases, it will show how, in a moment of uncertainty, our actions resolved a battle between our desires and subsequently altered our volitional makeups.¹⁹ In still other cases, including cases of *akrasia*, it will describe how our actions were out of sync with our central volitions. In many situations of each kind, we act responsibly and further the stories of our selves. We should resist the urge to restrict attributions of responsibility or meaningful self-creation to any one type of narrative.

¹⁹ Cf. Frankfurt's (1987) claim that deliberate decisions serve as guides for other decisions and thus tend to integrate us over time as people who care about the volitions that motivated the decisions (p. 175).

***Two Sorts of Self-Creation:
On Galen Strawson's "Basic Argument"***

Articulating Our Values

One way in which we can shape our wills, then, is through our decisions to act in particular ways. But we can also shape our wills more indirectly, and by laying out a model of the acting self in the previous section, I have given the context for a detailed account of this capacity. For I have suggested that our selves are constituted in part by a number of our values, concerns, and projects, not all of which mutually cohere. Further, our volitional makeups are not static; they alter as a consequence of our actions and our experiences over time. Charles Taylor (1976) supplies yet another complexity: when we attempt to articulate our most fundamental values—and so to unify many of the inconsistent volitions that are a part of our selves—it is precisely these fundamental values that are “least clear, least articulated, most easily subject to illusion and distortion” (p. 124). This means that any product of our attempts at articulation will not only be incomplete but may be quite wrong, and so the question of whether we understand ourselves rightly can always arise anew.

This inchoateness of our fundamental values makes possible indirect control over our wills. The formation of our wills is, as Frankfurt (1982) writes, a matter of coming to care about particular things rather than others, and we may indirectly affect this process by *caring about what we care about* (p. 91). This involves reflecting on the worthiness of the things we care about and, since our criteria for worthiness will necessarily be shaped by just these same cares, it also involves remaining open to different criteria than those we currently hold (cf. Taylor, 1976, pp. 124-5). In reflecting on our fundamental values, we neither choose them *ex nihilo* nor simply express the immutable, as Taylor (*ibid.*) notes:

Our attempts to formulate what we hold important must, like descriptions, strive to be faithful to something. But what they strive to be faithful to is not an independent object with a fixed degree and manner of evidence, but

LYCEUM

rather a largely inarticulated sense of what is of decisive importance. An articulation of this “object” tends to make it something different from what it was before. And by the same token a new articulation doesn’t leave its “object” evident or obscure to us in the same manner or degree as before. In the act of shaping it, it makes it accessible and/or inaccessible in new ways. (pp. 123-4)

In articulating our fundamental values, we do shape ourselves in genuinely new ways, but always attempting to be faithful to our own desires and loves, the values of those communities with which we identify, and our best understandings of what is good.

This self-shaping, obviously, does not involve the ability to make “bare choices.” In fact, the decisive importance we place on our fundamental values suggests that articulations of these values often constrain us from acting against them (though, of course, not in every case; *akrasia* often remains possible). This might seem like a limitation of our control over our selves rather than a manifestation of it. But, as Gary Watson (2002) reminds us, we should remember that “the assertion of moral or volitional necessity is not just the assertion of a constraint on one’s deliberative judgments . . . but part of the content of those judgments” (pp. 108-9)—that is, it expresses the vital importance of the judgment in question to us. We often do not consider this constraint on deliberation to rob us of responsibility; rather, as Frankfurt points out, we often take responsibility for our deep values, approving of our inability to go against what we take to be fundamentally important.²⁰

²⁰ Moreover, even if not everyone thinks worries about volitional necessity’s control-undermining tendencies can be so easily dispelled, those who take *causa sui* status with respect to one’s volitions to be necessary for responsibility for our actions should not be among those worried; after all, on their account, one’s actions are a function of one’s character traits, and this allows for ultimate responsibility for one’s actions (provided one is *causa sui*).

Two Sorts of Self-Creation: On Galen Strawson's "Basic Argument"

What about cases in which this volitional necessity *does* have the phenomenology of constraint—when we really wish we could seriously consider acting against our deep volitions? The standard example of volitional necessity in the literature is Martin Luther's (possibly apocryphal) closing words to the Diet of Worms: "Here I stand; I can do no other. God help me."²¹ But, as Nomy Arpaly (2006) wryly remarks, "one can feel chosen by God and yet cry, 'God! Can't you for once choose someone else?'" (p. 126). The constraints these volitions impose on our deliberation sometimes prevent us from performing other actions—perhaps more prudent or noble actions—that we simultaneously desire to perform.²²

But this framing exaggerates the extent of the problem by considering volitional necessity only from the perspective of the agent during deliberation. From a broader, narrational perspective, the import of necessity may be quite different. In some cases, we will ultimately come to identify with the constraint and embrace it, ceasing to reserve acceptance of our desires as they are.²³ In other cases, we may endure the constraint of cares and loves we do not endorse for long periods of time before finally prevailing and throwing them off. Taylor (1989) is right to insist that "the issue of [the human] condition can never be exhausted for us by what we *are*, because we are always also changing and *becoming*" (pp. 46-7, italics original). This is as true for our ability to shape ourselves as for any other aspect of our moral existence. And so, as long as the

²¹ I think the example was first used by Dennett (1984b).

²² This answers a question posed by one referee for this journal, namely, whether the sort of self-creation I argue that humans actually have is sufficient to dispel all moments of frustration at our lack of Strawsonian self-creation. This question should be answered in the negative, as the quotation from Arpaly above suggests. Nevertheless, as will be argued below, while it is at times frustrating to lack the capacity for something like Strawsonian self-creation, on balance, it would even be more frustrating to *possess* this capacity.

²³ See Frankfurt (1977, pp. 63-4). It is important to note that this occurs not only in cases where we finally accede to our desires while, in a significant sense, continuing to affirm our previous judgment that they were not part of our ideal selves, but also in cases where we accede to those desires precisely as we realize that our previous evaluations of them were mistaken.

LYCEUM

states of our wills confront us as problems, we can never rule out the possibility that our moments of impotence will find their places in narratives of developmental or redemptive self-shaping.

But some people never prevail against their truly bad volitions. So this modest sort of self-creation does not preclude our failing to achieve the goods we desire as moral agents. The question, then, is whether it achieves more of them than the ability to be *causa sui* (in, recall, our somewhat restricted sense). In the next section, I will argue that it does, and thus that it paradigmatically enables identification with self.

Causa Sui or Contextual Self-Creation?

Let's take a step back for a moment. What aspects of being *causa sui* seem desirable to us? The answer may seem obvious: it is the ability to *cause* our selves—that is, to have complete and direct control over them. But I think the correct answer goes deeper, for the desirability of causal influence in moral agency is derivative: causal influence correlates with intrinsically significant desiderata. What are these desiderata? They include our sense of responsibility for our wills and our reflective endorsement of our volitions. Some measure of causal influence over our wills is necessary for each of these desiderata; their absence can give rise to feelings of alienation from oneself.

We may call these desiderata *personal*, since they involve only facts about the agent's relationship to her will. But there are other important conditions for identifying with our wills.²⁴ Some of these are *interpersonal*: we want to have

²⁴ Below I discuss *interpersonal* and *suprapersonal* desiderata for moral agency. But Robert Bishop suggests that some desiderata for moral agency are also *impersonal*; these describe states of the mind-independent world that are necessary conditions for the possibility of moral agency. For instance, if universal determinism and causal closure obtain, then perhaps our qualitative experience is epiphenomenal and responsible agency is impossible. Even if this is correct, however, I think it is fitting that this category of desiderata receives less coverage than the other three, because the claim that there really are desiderata for moral agency assumes that the impersonal desiderata are met; if

Two Sorts of Self-Creation: On Galen Strawson's "Basic Argument"

values which some community or communities around us will share. And some might be called *suprapersonal*: many agents desire to value those things they *ought* to value, quite apart from considerations of what they or any communities around them actually endorse. Of course, we cannot step outside our present commitments or all the communities to which we belong, seeing bare ethical facts from “the view from nowhere.” But this doesn’t preclude conformity with objective normative truth from being a sensible regulative ideal (and thus a sensible desideratum).²⁵ Now, consider: does the bare self have the resources to enable us to achieve these suprapersonal goods? I think not, for an important and perhaps surprising reason: the suprapersonal goods depend strongly on the interpersonal goods—i.e. on acting intelligibly within the norms and values of communities to which the agent belongs—in three ways.

First, while there are truths about the manners of existence suitable for human beings that are not grounded in the beliefs of human beings (either individually or collectively),²⁶ this does not entail that either questions of the good life or their concrete answers can be formulated intelligibly without

they are not, we do not in fact have any desires at all. And if the impersonal desiderata are met, we do not need to act to meet them further; these desiderata do not obtain in degrees.

²⁵ As my framing of this point may suggest, I favor some version of ethical realism, but I don’t think the point depends on it: even if, say, moral expressivism were true, there seems no reason why there should not be norms governing what the person expresses in particular circumstances, quite apart from her beliefs or the beliefs of the surrounding communities about them. (That is, even if moral expressions have no truth conditions independent of the commitments of the speaker, their expression may be subject to prudential and epistemic norms.) Moreover, there is no reason in principle why these suprapersonal desiderata could not be constructed from intersubjective materials along quasi-realist lines.

²⁶ The independence in question concerns the *grounds* of these truths, not their *reference* or *formulation*. I am only arguing that even if all human beings informed one another about their contextual visions of the good and came to an agreement about “the good lives for human beings,” this formulation might still be incomplete and even mistaken. As will become clear, I am quite ready to grant that the meaningfulness of these truths depends on their translations into limited, contingent contexts. Further, these truths are, no doubt, multivalent; I am not suggesting that any claim as simple as, e.g., “the good human life is the life of study” is true or plausible.

LYCEUM

reference to a socio-historical context or the norms it espouses. As Charles Taylor (1989) remarks, questions like *who am I?* and *who should I be?* find their meaning for us in intersubjective dialogue (see p. 35). Further, to be a *good human agent* is, in part, to achieve excellence in various elements of one's social life, including comporting oneself with norms particular to one's communal situation. But this condition, of course, has no actual content when abstracted from all particular social contexts, and it is just these contexts from which the bare self is removed.

Now Galen Strawson might retort that this is precisely the problem: that these facts only highlight the all-encompassing influence of environment that gives rise to the problem of the Basic Argument in the first place. But Taylor (1989) disagrees, contending that the social embeddedness of questions of identity "cannot be just a contingent matter, [since] there is no way we could be inducted into personhood except by being initiated into a language" (p. 35). And language, in turn, cannot be divorced from its social context. So even if moral facts are not socially contextual, both moral language and moral life are—and so the *causa sui* self exists outside both.

Second, even granting that some moral facts are *eo ipso* binding reasons for all rational agents, it seems doubtful that all goods that strongly influence our moral agency can be translated into this "dialect of universal reason." Suppose a person is deeply concerned about which vocation will be hers. In virtue of the particular vocation she pursues and about which she is concerned, she will be drawn to behave in particular morally significant ways rather than others and to achieve particular moral goods rather than others. So the question of which vocation an agent will pursue is a question that is pivotal for her moral agency. But no plausible universal moral law will dictate that she ought to pursue any particular vocation. This is why even Kant, who argued for as strong an association between the dictates of morality and those of universal reason as anyone, recognized (roughly) that moral agents require contingent desires to fill in the gaps left by the moral law, if they are to get around to any concrete moral

Two Sorts of Self-Creation: On Galen Strawson's "Basic Argument"

action.²⁷ But as the bare self is removed from the grip of all contingent desire, it loses the ability to determine its will sensibly toward particular moral actions.

Third, not only does the performance of moral actions depend on contingent particularities, but the *moral character* of moral actions is similarly context-dependent. For if one lacks all such particular interests and cares, one cannot pursue any good with the affective dispositions necessary for genuine moral action.

Here a less abstract formulation may help. Marriage is a form of community that normatively carries with it particular moral goods; people who are married ought to be loyal, self-sacrificial and encouraging toward one another in distinctive ways. Presumably, *ex hypothesi*, the bare self recognizes these moral features of marriage; let us assume (somewhat fatuously) that, in virtue of these features, it decides to marry. How would it decide whom to marry? Stripped of contingent desires and contexts, it could do so only by considering the magnitude of the moral goods to be achieved by marrying any particular person (i.e., who would benefit the most from its loyalty, who would benefit *it* the most by their self-sacrifice, etc.). But I submit that this sort of calculation simply is not a morally acceptable basis for making this decision—not to mention that, in practice, it seems questionable if such considerations could actually sufficiently motivate even a *causa sui* human being to behave morally in these ways over a long period of time. The moral basis for devoting oneself to marriage, I argue, is affection for one's partner *considered as a particular*, and while this affection must be sensitive to the (morally salient) universals involved, it cannot be exhaustively expressed by reference to them. And, I submit, similar remarks apply to the selection of our vocation, our

²⁷ See *Religion Within the Bounds of Mere Reason*, Bk. I. More precisely, Kant recognizes that human beings are necessarily motivated by their desires and requires only that moral agents subordinate their self-serving desires to their desire to follow the moral law for its own sake. There is never any question of the moral agent's quelling all "non-dutiful" desires within her, since this is neither possible nor productive for human beings. Indeed, the very concept of an imperfect duty makes clear the great role that contingent experiences and desires have to play in moral agency.

LYCEUM

interests, the causes important to us, and so on. Moral agency, then, is motivated by loves with irreducibly particular dimensions. But our loves for bare particulars must be fixed by the contingent contexts in which we are embedded.

In summary, the question of the good *for us*—us contingent, contextually limited human agents—cannot be simply a question of the Abstract, Timeless Good. Taylor (1989) perceptively asserts that “our orientation in relation to the good requires not only some framework(s) which defines the shape of the qualitatively higher but also a sense of where we stand in relation to this” (p. 42). The arguments above show that even if we could consider outside our limitations the question of where we should stand, we would not be able to find an answer. For any space in which we could place ourselves would fall within the limitations we would have transcended, and selection between such spaces is possible only by entering into them. The question of where we stand with respect to that which is valuable is answerable only if we are given somewhere to stand, helped (and *made*) to begin moving, and finally prompted to select—increasingly, though never fully, on our own—between the paths that confront us. By preventing us from beginning in interpersonal space, total self-creation precludes our attaining not only the interpersonal desiderata of agency but also the suprapersonal ones.

So, even if total self-creation allows us more responsibility for ourselves and thus achieves the personal desiderata of agency better than modest self-creation, I argue that it prevents us from achieving goods at least as vital (and probably more so). But I doubt that even this much can be said for total self-creation. For whatever sorts of control over oneself being *causa sui* may involve, it structurally rules out being influenced by one’s context in ways that aren’t already latent in one’s prior self. And, insofar as it stipulatively rules out one’s acting in ways that are not determined by one’s deep character, it precludes not only one’s reformulation of oneself but also one’s coming to identify with oneself through protracted struggle against *akrasia*. But if the self one has is not rendered intelligible by any particular context or location, and if one does not come to identify with it through genuine development or through struggling to

Two Sorts of Self-Creation: On Galen Strawson's "Basic Argument"

express one's fundamental values in action—in short, if one does not have a history, a narrative by which one becomes intelligible to oneself—how could one identify with or take responsibility for oneself, the strong sort of control one has over oneself notwithstanding? How could one fail to experience alienation? Taylor (1989) is right to claim that “to ask what a person is, in abstraction from his or her self-interpretations, is to ask a fundamentally misguided question” (p. 34). But it is one's *story*, not one's control over oneself, that allows one to interpret oneself.

Conclusion

If the argument of the preceding section is correct, then we have strong evidence against Galen Strawson's claim that the understanding of responsibility for self found in the Basic Argument is central to ordinary thought about morality and justice: namely, that only self-creation of a very different sort can provide us with the goods we seek in moral agency. We should not be distressed, then, if Strawson's ultimate responsibility is unattainable, for the self-creation that mature humans practice better allows for meaningful agency and identification with self.

The argument of this paper, then, provides libertarians and compatibilists alike with the necessary resources to found a defense against a prominent skeptical argument. But I think it does this by stepping back and considering the points that are indispensable to any account, libertarian or compatibilist, that would accurately describe the agency we value—namely, that meaningful agency develops out of the influences of heredity, environment and community; through the interplay of volitions and actions; and by our gradually coming to take responsibility for our selves and to reflectively articulate our senses of what is fundamentally valuable. If a theory of free will, from whichever side of the

LYCEUM

divide, takes these points as pivotal, it will not fail to capture much that is deeply important concerning human life and activity.²⁸

References

Arpaly, Nomy. (2006). *Merit, meaning, and human bondage: An essay on free will*. Princeton, NJ: Princeton University Press.

Dennett, Daniel C. (1984a). *Elbow room: The varieties of free will worth wanting*. Cambridge, MA: The MIT Press.

———. (1984b). “I could not have done otherwise—so what?” Edited version reprinted in Kane (2002a), pp. 83-94.

Fischer, John Martin. (1999). “Responsibility and self-expression.” Reprinted in Fischer (2006), pp. 106-23.

———. (2005). “Free will, death, and immortality: The role of narrative.” Reprinted in Fischer (2009), pp. 145-64.

———. (2006). *My way: Essays on moral responsibility*. New York: Oxford University Press.

———. (2009). *Our stories: Essays on life, death, and free will*. New York: Oxford University Press.

Frankfurt, Harry G. (1977). “Identification and externality.” Reprinted in Frankfurt (1998), pp. 58-68.

²⁸ My thanks are due to Mark Talbot, Robert Bishop, Jay Wood, two anonymous referees for this journal, and especially Stephen Scheidell for helpful comments and suggestions.

***Two Sorts of Self-Creation:
On Galen Strawson's "Basic Argument"***

- . (1982). "The importance of what we care about." Reprinted in Frankfurt (1998), pp. 80-94.
- . (1987). "Identification and wholeheartedness." Reprinted in Frankfurt (1998), pp. 159-76.
- . (1998). *The importance of what we care about*. New York: Cambridge University Press.
- Ginet, Carl. (1989). "Reasons explanation of action: An incompatibilist account." *Philosophical Perspectives* 3, pp. 17-46.
- Hobart, R. E. (1934). "Free will as involving determination and inconceivable without it." *Mind* 43(169), pp. 1-27.
- Kane, Robert. ed. (2002a). *Free will*. Malden, MA: Blackwell Publishers.
- . (2002b). *The Oxford handbook of free will*, 1st ed. New York: Oxford University Press.
- Nagel, Thomas. (1986). *The view from nowhere*. New York: Oxford University Press.
- Strawson, Galen. (1994). "The impossibility of moral responsibility." Reprinted in Watson (2003), pp. 212-28.
- . (2000). "The unhelpfulness of indeterminism." *Philosophy and Phenomenological Research* 60(1), pp. 149-55.
- . (2002). "The bounds of freedom." In Kane (2002b), pp. 441-60.

LYCEUM

- . (2004). “Free agents.” Reprinted in G. Strawson (2008), pp. 359-86.
- . (2008). *Real materialism and other essays*. New York: Oxford University Press.
- Taylor, Charles. (1976). “Responsibility for self.” Reprinted in Watson (1982), pp. 111-26.
- . (1989). *Sources of the self*. Cambridge, MA: Harvard University Press.
- Watson, Gary. (2002). “Volitional necessities.” Reprinted in Watson (2004), pp. 88-122.
- . Ed. (2003). *Free will*. 2nd ed. New York: Oxford University Press.
- . (2004). *Agency and answerability: Selected essays*. New York: Oxford University Press.

About the Authors

David Cheely is an adjunct instructor in the Honors College at the University of South Florida (USF). Originally from Michigan, he earned a BA in Philosophy from Michigan State University, an MA in Philosophy from USF, and is currently working on his PhD at USF. His dissertation focuses upon the socio-political ramifications of narrative approaches to Self Identity interpreted through the theoretical framework of Martin Heidegger.

Leonard Ferry is an ABD PhD candidate at the University of Toronto with research interests that focus on the thought of St. Thomas Aquinas. In addition to a book that he co-edited with Rebecca Kingston, *Bringing the Passions Back In* (University of British Columbia Press, 2008), he has published essays on virtue ethics, the natural law, and Aquinas' anthropology and emotion theory in journals including *Maritain Studies*, *Logos: A Journal of Catholic Thought and Culture*, *New Blackfriars*, and *The Heythrop Journal*.

Sangkyu Shin is assistant professor of Ewha Institute for the Humanities at Ewha Womans University. He received his PhD at the University of Texas at Austin in 2003 with the thesis, *Meaning and Normativity – A Study of Teleosemantics*, which focuses on the problem of naturalizing intentionality. He is the author of *Artificial Intelligence and the Future Condition of Human Being* (in Korean). He has published articles (in Korean) on intentionality, phenomenal consciousness, the extended mind, and transhumanism.

Griffin Klemick earned his BA in philosophy from Wheaton College (IL) and is currently an MA student in philosophy at Northern Illinois University. He is interested in community, practice, and understandings of what is valuable, because he thinks these are central to, and inextricable from, any plausible answer to any deep philosophical problem. He likes writing on “big questions” concerning mind and world, as well as ethics and philosophy of action.