ABSTRACT

THE DEEP STRUCTURE OF LAW AND MORALITY

This Article argues that morality and law share a deep and pervasive structure, an analogue of what Noam Chomsky calls the "deep structure" of language. This structure arises not to resolve linguistic problems of generativity, but rather from the fact that morality and law engage psychological adaptations with the same natural function: to allow us to resolve social contract problems flexibly. Drawing on and extending a number of contemporary insights from evolutionary psychology and evolutionary game theory, this Article argues that we resolve these problems by employing a particular class of psychological attitudes, which are neither simply belief-like states nor simply desire-like states, though they bear affinities to both. The attitudes are "obligata."

Obligata breathe life into our moral and legal practices, and have a specific structure. They blend agent-centered attitudes toward persons with attitudes toward shared standards for action as producing reasons that exclude some arising from personal interest. Obligata are "judgment-sensitive attitudes": reasons can be sensibly asked and offered for them. They incline us to react critically to deviations and perceive these reactions as warranted. Obligata nevertheless sensitize us to the standard excuses, thereby allowing us to mend our relationships after some seeming breaches. We express obligata in the special normative terminology that morality and law share, including in contexts of discussion and dispute that can become incredibly charged. In these interactions, obligata allow us to meaningfully disagree, and sometimes thereby reach consensus, even when our resolutions are not traceable to any particular reasons we antecedently accepted. This talk thus engages underlying psychosocial mechanisms that can—in the appropriate social and political circumstances—help us maintain sufficient agreement over what we owe to one another to live well together. Obligata thereby allow us to *enjoy* our lives together. Finally, it is possible that our moral and legal judgments supervene on natural facts because there are natural facts—about what moral and legal rules would conduce to all our objective individual interests in the right way—that partly explain the shape that morality and law take in our lives.

The structure of obligata *is* the deep structure of morality and law. This suggests that much of the legal literature—including familiar descriptive and normative accounts from law and economics scholars—have been presupposing a psychological picture that is deeply at odds with how we naturally think about obligation. Morality and law do not arise from, and could not be sustained only by, separable beliefs about the world and preferences for states of affairs. The challenge raised here runs deeper, however, than recent empirical work showing we deviate from instrumental rationality in numerous, systematic ways. Our capacities to reason instrumentally may not figure very centrally at all in our moral or legal practices, and we may necessarily misunderstand these normative phenomena if we keep trying to shoehorn them into that model. To understand morality and law, we must instead understand how our distinctive capacities to identify and respond appropriately to obligations function.

THE DEEP STRUCTURE OF LAW AND MORALITY

by

ROBIN BRADLEY KAR*

Even the casual observer will note that law and morality resemble one another in numerous and striking ways. Both practices typically consist of rules with general applicability, which we perceive to have special importance in our lives and to provide us with personal mandates that can operate irrespective of at least some consequence. Both purport to provide us with reasons to act that can override other compelling ones that arise from personal interest. Both also contain a special normative vocabulary—including terms like "ought," "duty," "obligation," "excuse," "right," and the like[1]—which terms are essentially contestable[2] and irreducible in meaning to any descriptive statements of natural fact.[3] Yet we tend to believe that there can be no

* Professor Kar is an Associate Professor of Jurisprudence and Law and Loyola Law School in Los Angeles, and Deputy Director of their Center for Interdisciplinary and Comparative Jurisprudence. He is a graduate of Harvard University and Yale Law School, and received his doctorate in moral and legal philosophy from the University of Michigan. This piece has profited enormously from comments and conversations—in many cases, spanning years—with Elizabeth Anderson, Ian Ayres, Jules Coleman, Stephen Darwall, Brianna Fuller, Allan Gibbard, Brian Leiter, Peter Railton, David Velleman, Ken Walton and Lauren Willis. As is always the case in these matters, all remaining errors are mine. Many thanks are also in order to the Woodrow Wilson Foundation, which funded the final stages of this project through a Charlotte W. Newcombe Fellowship, and to the Horace H. Rackham School of Graduate Studies, which funded earlier parts through a Rackham Predoctoral Fellowship.

[1] *See, e.g.*, JOSEPH RAZ, *Introduction*, *in* THE AUTHORITY OF LAW vii (1979). Joseph Raz notes that this characteristic of legal language is sometimes thought to support the natural law thesis that law is "inescapably moral." *Id.* Neither Raz nor I take this conclusion to be inevitable.

[2] To say that a term is "essentially contestable" is to say that whatever claims one makes using the term can be meaningfully debated; terms like these are thus "variously describable" and typically "admit[ ] of considerable modification in the light of changing circumstances." W.B. Gallie, *Essentially Contested Concepts*, 56 PROC. ARISTOTELIAN SOC. 167, 167-78 (1956) (noting that many of our moral and political concepts are essentially contestable). Ronald Dworkin has argued that a number of features of our legal discourse in contexts of adjudication establish that our legal concepts are also essentially contestable. *See, e.g.*, Ronald Dworkin, *Hard Cases*, *in* TAKING RIGHTS SERIOUSLY 103 (1977).

[3] The classic source of this insight is G.E. MOORE, PRINCIPIA ETHICA, 6-20 (1971). Moore charged any attempt to define the "good" in purely naturalistic terms as falling prey to a "naturalistic fallacy," and clarified ways in which the meaning of this normative concept appear irreducible to any empirically definable concept. *Id.* W.D. Ross later extended this form of argument to deontological concepts like

warranted difference in legal or moral judgment without some difference in the natural facts.[4] And our moral and legal practices are pervaded, in similar ways, by standards that not only purport to provide us with reasons to act but also to criticize deviations in ways that imply the permissibility of certain forms of sanction or coercion[5]—as well as by a portfolio of standard excuses that operate to defeat such criticisms in strikingly similar ways.[6]

This Article argues that these resemblances are more than superficial. They arise from the fact that law and morality share a deep and pervasive structure, an analogue in the moral and legal domain of what Noam Chomsky has called the "deep structure" or "universal grammar" of language.[7] This structure arises from the fact that morality and law engage psychological adaptations with the same natural function: to allow us to resolve various classes of social contract problems[8] flexibly. Drawing on and extending a number of contemporary insights from evolutionary psychology and

---

"right" and "duty," and charged Moore with falling into a similar fallacy when trying to define the right in terms of the good. *See* W.D. ROSS, THE RIGHT AND THE GOOD 8-11 (1930).

[4] The technical term for this is to say that these normative terms "supervene" on natural facts. To say that a moral property "supervenes" on a non-moral property is to say that two items cannot differ in their moral properties without differing in some non-moral property as well. *See, e.g.*, R.M. HARE, THE LANGUAGE OF MORALS 80 (1952). It is common to observe that moral properties apparently supervene on natural facts. *See, e.g.*, Michael Ridge, *Moral Non-Naturalism* § 6, *in* STANFORD ENCYCLOPEDIA OF PHILOSOPHY, http://plato.stanford.edu/entries/moral-non-naturalism; Michael S. Moore, *Moral Reality Revisited*, 90 MICH. L. REV. 2424, 2515-17 (1992). Our legal judgments also apparently supervene on natural facts, as is reflected in the thought that "like cases should be treated alike" under the law. *See, e.g.*, John E. MacKinnon, *Law and Tenderness in Bernhard Schlink's* The Reader, 16 CARDOZO STUD. L. & LIT. 179, 187 (2004) (observing this link and defining "legal supervenience" as the claim that "[i]f 'all things,' or relevant features, are equal, . . . then the legal character that attaches to that range of features must be 'the same' in all cases where that range occurs."); Larry Alexander & Ken Kress, *Against Legal Principles*, 82 IOWA L. REV. 739, 764 n.96 (1997) (distinguishing ways that legal and moral judgments supervene on natural facts, but observing that legal judgments supervene in a complex way).

[5] *See, e.g.*, H.L.A. HART, THE CONCEPT OF LAW 82-91 (2d ed. 1961) (making this point in relation to morality and law); RICHARD B. BRANDT, A THEORY OF THE GOOD AND THE RIGHT 163-70 (1979) (making point in relation to morality).

[6] *See generally* H.L.A. Hart, *Legal Responsibility and Excuses*, *in* DETERMINISM AND FREEDOM IN THE AGE OF MODERN SCIENCE 81 (Sidney Hook ed., 1958), *reprinted in* PUNISHMENT AND RESPONSIBILITY 28, 29-53 (1968) (noting pervasiveness of certain specific excuses in different areas of the law); Richard B. Brandt, *A Utilitarian Theory of Excuses*, 78 PHIL. REV. 337 (1969), *reprinted in* RICHARD B. BRANDT, MORALITY, UTILITARIANISM, AND RIGHTS 215-234 (1992) (noting pervasiveness of these excuses in morality); Richard B. Brandt, *A Motivational Theory of Excuses in the Criminal Law*, *in* 27 NOMOS: CRIMINAL JUSTICE 165 (J. Roland Pennock & John W. Chapman eds., 1983), *reprinted in* BRANDT, *supra*, at 235-262.

[7] *See, e.g.*, NOAM CHOMSKY, REFLECTIONS ON LANGUAGE 29-30 (1975) (defining "universal grammar"); NOAM CHOMSKY, ASPECTS OF THE THEORY OF SYNTAX 136 (1965) (using terminology of "deep structure" to refer to this phenomenon).

[8] The term "social contract problem" will be used most generally to refer most generally to any situation in which all parties beginning from a suitably defined starting position would agree to be bound by certain rules on the condition that all others would be similarly bound. This general definition admits of a number of alternative specifications, and the precise sense in which our moral and legal psychologies function to resolve social contract problems will be elaborated further in the course of this Article.

evolutionary game theory, this Article will develop the claim that we resolve these problems by employing a particular class of psychological attitudes, which are neither simply belief-like states nor simply desire-like states, though they bear affinities to both. The attitudes will be called "obligata," for reasons to be explained.[9] As they appear in us, they are a peculiar blend of propositional attitudes,[10] deontological motivations to follow rules, and reciprocally conditioned expectations of and attitudes toward other persons. Obligata are also judgment-sensitive attitudes—in the sense that reasons can be sensibly asked or offered for them[11]—and they are bound up with a number of motives and familiar moral emotions, like shame and guilt.[12] They are the attitudes we express when we engage in moral and legal discussion.

Obligata constitute our sense of obligation, and thereby breathe life into our moral and legal practices. Their structure *is* the deep structure of law and morality.

An understanding of obligata will, moreover, have important consequences not only for legal theory but also for how we should approach normative proposals in law. There is a well-developed and long-standing strain of scholarly literature—predominantly, though not only, arising in the law and economics movement—that either explicitly or implicitly presupposes a very different psychological picture of us as acting primarily on the basis of separable beliefs about the world and desires (or preferences) for various states of affairs.[13] On this common view, our practical reasoning is purely instrumental, and proponents of this view sometimes claim that we only have individual reasons to pursue things like our considered preferences. More recently, a number of researchers have begun to document numerous ways that we in fact deviate from this so-called *Homo Economicus* model, and have made efforts to accommodate the fact that we sometimes apparently exhibit desires that are altruistic or

---

[9] *See* Section C, *infra*.

[10] A "propositional attitude" is a relational mental state that connects a person to a proposition, or to what is asserted in uttering a sentence. Propositions are often thought of as the simplest components of thought and as expressive of meanings, or contents, that can be true or false. There are a number of common attitudes that we can have to a given proposition. We can *believe* the proposition; *hope* for it; *wonder* about it; *expect* it; and so on. All of these would thus be propositional attitudes on the orthodox definition.

[11] The term "judgment-sensitive attitude" is used in the sense that T.M. Scanlon has made familiar. Scanlon defines "judgment-sensitive attitudes," with innocent circularity for present purposes, as the class of attitudes that "an ideally rational person would come to have whenever that person judged there to be sufficient reasons for them and that would, in an ideally rational person, 'extinguish' when the person judged them not to be supported by reasons of the appropriate kind." T.M. SCANLON, WHAT WE OWE TO EACH OTHER 20 (2000). Our judgments that something is right or wrong, or is required or prohibited by law, are sensitive to reasons in this sense.

[12] For a description of the link between moral emotions like shame and guilt to our normative practices, see Daniel M.T. Fessler & Kevin J. Haley, *The Strategy of Affect: Emotions in Human Cooperation*, *in* GENETIC AND CULTURAL EVOLUTION OF COOPERATION 7-37 (Peter Hammerstein ed., 2003); Jonathan Haidt, *The Moral Emotions*, *in* HANDBOOK OF AFFECTIVE SCIENCES 852-870 (R.J. Davidson et al. eds., 2003).

[13] *See, e.g.* Dan M. Kahan, *Trust, Collective Action, and Law*, 81 B.U.L. REV. 333, 333 (describing this model as providing the basis for the "standard tropes of public law").

other-regarding.[14] The challenge posed in this Article will, however, run deeper. It will suggest that the basic social psychological building blocks out of which we create and sustain our moral and legal relations have, and must have, a deep structure that is fundamentally at odds with current economic frameworks.[15] If this is correct, then much of the scholarly literature has been presupposing a psychological picture of us that is importantly incomplete, and the contours of which play a much smaller role in morality and law that has often been assumed. If we hope to approach normative questions about the law from the right angle and with the right clarity, we must therefore learn to understand better how our natural sense of obligation functions.

A. PRELIMINARY CLARIFICATIONS[16]

As indicated, the main portions of this Article will argue that law and morality share a deep and pervasive structure and will elaborate what, precisely, that structure consists in. Before turning to that project, a number of preliminary clarifications are in order.

As an initial matter, the domain of morality at issue in this Article is narrower than what some people mean by the term. In ordinary speech, people sometimes use the term "moral criticism" to refer to a wide range of objections that people sometimes raise to various forms of behavior or conduct.[17] As T.M Scanlon has observed, "[v]arious forms of behavior, such as premarital sex, homosexuality, idleness, and wastefulness, are often considered immoral even when they do not harm other people or violate any duties to them."[18] When people engage in this kind of criticism, they do not, however, typically believe that people have *obligations* to others with respect to these activities. To think that one has an *obligation* to perform an act is to think the following four things: that (i) there is a standard of action that has some general applicability, or applicability irrespective of the antecedent desires or interests of the persons to whom it applies; (ii) the standard provides each person to whom it applies with a reason to act, irrespective of her antecedent desires and interests; (iii) the standard provides each person to whom it applies with a reason to act that overrides or excludes[19]

---

[14] *See, e.g.*, THE LAW AND ECONOMICS OF IRRATIONAL BEHAVIOR (Franceso Parisi & Vernon L. Smith eds., 2005); Elizabeth Anderson, *Beyond* Homo Economicus, 29 PHIL. & PUB. AFF 170-200 (2000) (discussing empirical evidence and modern attempts to account for some of it from within basic economic frameworks).

[15] For another set of reasons to question whether our commitments to social norms can be understood instrumentally, see Anderson, *supra* note **Error! Bookmark not defined.**, 170-200.

[16] This section has profited from many helpful conversations with Brianna Fuller, who will—I sincerely hope—join us in legal academia one day.

[17] *See generally* SCANLON, *supra* note 11, at 6.

[18] *Id.* Neither Scanlon nor I are committed to the view that objections of this kind are necessarily merited or morally serious.

[19] The term "exclusionary reason" is used here in the sense made familiar by Joseph Raz. *See, e.g.*, JOSEPH RAZ, PRACTICAL REASON AND NORMS 35-45 (1990). This use does not entail commitment to Raz's particular way of categorizing first and second order reasons, or to his idea that moral reasons are first order reasons that can be silenced by legally authoritative reasons.

other compelling ones arising from antecedent desire or interest;[20] and (iv) breach of the standard gives some other person or group standing to complain and/or warrants what would otherwise be resented, namely, certain forms of punishment or coercion for non-compliance.[21] Scanlon has usefully referred to this part of morality as "what we owe to each other," and has suggested that "this part of morality comprises a distinct subject matter, unified by a single manner of reasoning and by a common motivational basis."[22]

The law similarly purports to provide us with obligations that meet the four criteria under discussion. Indeed, when a real or alleged political authority attempts to induce people to act by means of imperatives backed by force, those imperatives will only be considered *legal* imperatives if they meet the same four criteria.[23] The domain of obligation is therefore the place where morality and law most plausibly intersect,[24] and this area of intersection will be the sole focus of this Article. The central thesis of this

---

[20] For a useful discussion of how these three characteristics are needed to give sense to the peculiar normative force (or 'categoricity') that moral obligations purport to have, see David Brink, *Kantian Rationalism: Inescapability, Authority and Supremacy*, *in* ETHICS AND PRACTICAL REASON 255-67, 280-87 (Garrett Cullity & Berys Gaut eds., 1997). Brink labels these three characteristics of moral obligations their (i) "inescapability"; their (ii) "authority"; and their (iii) "supremacy" (in the sense that moral reasons purportedly provide us with overriding reasons). *See id.* at 255. I have, however, replaced Brink's notion of "supremacy" with Raz's notion of an "exclusionary reason" because strict overridingness is not needed to account for the normative force of obligations in the broadest sense, and Raz's notion can capture the full range of possible obligations—whether moral, legal or otherwise—that we perceive ourselves to be under. *See generally* Joseph Raz, *Legitimate Authority*, *reprinted in* THE AUTHORITY OF LAW, *supra* note 1, at 16-25 (accounting for legal obligations in terms of exclusionary reasons).

[21] Stephen Darwall has recently emphasized that we cannot even begin to understand the distinction between the normativity of obligations and those of other purportedly categorical requirements, such as the requirements of logic, without conceding an intrinsic relation between obligations and others' standing to raise claims against one another. *See* Stephen Darwall, "*Because I Want It*," 18 SOC. PHIL. & POL'Y 129, 136-38, 144-53 (2001). Because morality and law purport to provide us with obligations, and not only categorical requirements like those of logic, this fourth criterion is needed to complete Brink's analysis. For an excellent early description of moral and legal obligation that is prescient in its sensitivity to this fact, see HART, *supra* note 5, at 82-91.

[22] SCANLON, *supra* note 11 at 6-7. I am, however, using the term "obligation" more broadly than either Scanlon or early Hart to refer to this whole domain. Scanlon uses the term "obligation" to refer only to those things that we owe to each other "arising from specific actions or undertakings." *Id.* Hart once limited his use of the term "obligation" to refer to those duties that "may be voluntarily incurred or created, . . . are owed to special persons (who have rights), [and] do not arise out of the character of the actions which are obligatory but out of the relationship of the parties." H.L.A. Hart, *Are There Any Natural Rights?*, 64 PHIL. REV. 175, 179 n.7 (1955).

[23] For detailed discussion of these points, see HART, *supra* note 5, at 18-25.

[24] Bernard Williams has, for example, famously argued that morality employs a special notion of obligation and gives obligations a special significance in deliberation. *See* BERNARD WILLIAMS, ETHICS AND THE LIMITS OF PHILOSOPHY 174-75 (1985). For Williams, this is a problem: He complains that morality turns everything into obligations, and considers deliberation to issue only in requirements and permissions, thus leaving out of the picture many other important dimensions of value. *Id.* at 175, 180-96. The same "criticism" would apply to the law, however, and Williams's criticism can thus be understood as suggesting that morality contains only a legalized conception of obligation. Be this as it may, this is the domain of morality that most plausibly intersects with the law, and this area of intersection will thus be the sole focus of this Article.

Article can thus be viewed as answering an important question in moral and legal psychology, namely: what psychological capacities must we have in order to identify and respond appropriately to obligations? The claim will be that the relevant capacities have a number of structural features, which will be elaborated.

A second important clarification relates the precise analogy that this Article will claim with Chomsky's work. Chomsky has argued that beneath the seemingly infinite variety of languages and linguistic expressions, there is a simpler set of rules[25]—which he calls "Universal Grammar"—that give us the capacity to identify and respond appropriately to language and to embed propositions into an indefinitely complex set of thoughts and statements.[26] Ever since Chomsky published his now familiar views, it has become fashionable—perhaps all too fashionable—to claim that there are "deep structures" to all kinds of complex social phenomena.[27] It is, however, important to take precaution before claiming any such extensions. Chomsky's claims about language find support in a constellation of distinctive considerations, which together reciprocally reinforce one another and provide his views with a particularly firm foundation. It is, moreover, the juxtaposition of these considerations that gives content to his precise notion that the features he identifies are "structural"—or innate and universal conditions of our capacities for natural language.[28] The most important of these bases are the following:

1. *Universality*. The linguistic phenomena that Chomsky calls "Universal Grammar" appear in all human languages, as disparate as many of them are on the surface, and as unrelated as they sometimes are by history or common ancestry.[29]

2. *Developmental Psychology*. Work in developmental psychology suggests that children develop linguistic competence in invariant stages and at relatively

---

[25] Chomsky may be using the term "rule" metaphorically when he talks about these phenomena as rules. What he is describing are conditions of linguistic intelligibility for us, at least if we are to use language to with its distinctive range of expression, but these conditions may not actually be "rules" in any ordinary sense of the word.

[26] *See, e.g.*, CHOMSKY, REFLECTIONS ON LANGUAGE, *supra* note 7, at 29-35; CHOMSKY, ASPECTS OF A THEORY OF SYNTAX, *supra* note 7, at 136-42.

[27] George P. Fletcher, for example, observed as early as 1997 that "a LEXIS check . . . revealed 247 usages of the phrase 'deep structure,' some in unexpected contexts." George P. Fletcher, *What Law Is Like*, 50 SMU L. REV. 1599, 1604 n.22. He points to the law review article *The Deep Structure of Capital Gains* as an example of such an unexpected context. *Id.*

[28] *See, e.g.*, CHOMSKY, REFLECTIONS ON LANGUAGE, *supra* note 7, at 29 ("Let us define 'universal grammar' . . . as the system of principles, conditions, and rules that are elements of properties of all human languages not merely by accident but by [biological] necessity . . . Thus [universal grammar] can be taken as expressing 'the essence of human language.' [Universal grammar] will be invariant among humans. [Universal grammar] will specify what language learning must achieve, if it takes place successfully. . . .What is learned must have the properties of [universal grammar], though it will have other properties as well, accidental properties. Each human language will conform to [universal grammar]; language will differ in other, accidental properties.").

[29] *See id.*

predictable times.[30] The capacities typically develop in ways that outstrip in complexity anything that the child could have learned solely from external sources of input or experience, at the relevant stages in his or her development.[31]

3. *Other Psychological Evidence*. Our linguistic competence can sometimes malfunction, even while many of our other cognitive and psychological capacities continue to operate perfectly well.[32] This fact could be explained if we had a specific adaptive capacity for language, which can selectively malfunction.[33]

4. *Functional Considerations: Generativity*. Our linguistic capacities allow us to express and understand a seemingly limitless number of thoughts, in part by embedding simpler thoughts into more complex linguistic or syntactic structures.[34] Chomsky has argued that in order to explain this feature of language, we must posit a more basic set of rules or conventions that allow us to combine simpler thoughts into more complex ones in specific ways. In technical jargon, we must view our linguistic capacities as "discrete combinatorial systems," which use a set of basic grammatical rules to allow us to express and understand the broad range of thoughts that we find expressible

---

[30] *See, e.g.*, STEVEN PINKER, THE LANGUAGE INSTINCT 32-45, 262-96 (collecting evidence from studies of language development in children that children reinvent the rules of generative grammar at typical stages in their development).

[31] *See, e.g.*, *id.* at 39-45, 276-96; CHOMSKY, REFLECTIONS ON LANGUAGE, *supra* note 7, at 4 ("A human language is a system of remarkable complexity. To come to know a human language would be an extraordinary intellectual achievement for a creature not specifically designed to accomplish this task. A normal child acquires this knowledge on relatively slight exposure and without specific training. He can then quite effortlessly make use of an intricate structure of specific rules and guiding principles to convey his thoughts and feelings to others, arousing in them novel ideas and subtle perceptions and judgments. For the conscious mind, not specifically designed for the purpose, it remains a distant goal to reconstruct and comprehend what the child has done intuitively and with minimal effort."); CHOMSKY, ASPECTS OF THE THEORY OF SYNTAX, *supra* note 7, at 27 ("A theory of linguistic structure . . . attributes tacit knowledge of [linguistic] universals to the child. It proposes, then, that the child approaches the data with the presumption that they are drawn from a language of a certain antecedently well-defined type, his problem being to determine which of the (humanly) possible language is that of the community in which he is placed. Language learning would be impossible unless this were the case."). This form of argument has come to be known as argument from "poverty of the stimulus." *See, e.g.*, *Editors' Introduction in* NOAM CHOMSKY, ON NATURE AND LANGUAGE 5-6.

[32] *See, e.g.*, PINKER, *supra* note 30, at 45 53, 297-331 (collecting evidence of this phenomenon). For example, patients who have been injured in Broca's area of the brain are unable to distinguish between grammatical and ungrammatical sentences, speaking instead in stunted phrases and strings of words. The opposite is true of those injured in Wernicke's area of the brain: they speak in grammatical sentences but produce nonsense, having lost the ability to retrieve words from their mental dictionary. *See id.* at 309-13.

[33] *See id.* at 299, 313.

[34] *See, e.g.*, CHOMSKY, ASPECTS OF A THEORY OF SYNTAX, *supra* note 7, at 136-42; CHOMSKY, REFLECTIONS ON LANGUAGE *supra* note 7, at 29-35; PINKER, *supra* note 30, at 84-125 .

in all natural languages.[35] This aspect of language is sometimes referred as its "generativity."[36]

How far might one go in finding analogies in morality and law for these four classes of considerations? As discussed more fully below, the literature now contains useful and suggestive work that one might try to analogize to the first three classes of considerations, but not the fourth. The fourth set is, however, absolutely critical to project of distinguishing genuinely structural features of our psychologies and lives from features that are merely accidental, even if universal or common.[37] In what follows, this section will therefore indicate what the current state of our knowledge is with regard to analogues of the first three classes of considerations. It will then discuss why there is no direct analogy to be made with regard to the fourth and clarify the implications of this fact for morality and law. It will then propose an alternative functional problem that morality and law solve that can replace language's generativity in developing a plausible account of the deep structure of morality and law.

Let us begin, then, with the first class of considerations and ask whether our moral and legal practices exhibit any universalities of the relevant kind. A number of people have tried to identify such universalities,[38] and one particularly probing and representative account is due to Donald E. Brown.[39] Inspired by Chomsky's conception of Universal Grammar, Brown has combed the wider ethnographic record looking for other culturally universal patterns in human life. Among those he claims to find are some that should be suggestive for moral and legal theory: giving, lending, possession, sense of self versus other, responsibility, voluntary versus involuntary behavior, intention, empathy, living in groups (which claim a territory and have a sense of being a distinct people), status and prestige (both assigned—by kinship, age, sex—and achieved), exchanges of labor, goods, and services, reciprocity (including retaliation), gifts, social reasoning, coalitions, government (in the sense of binding collective decisions about public affairs), authoritative rules, rights and obligation (including authoritative rules

---

[35] PINKER, *supra* note 30, at 84-88.

[36] *See, e.g.*, DONALD LORITZ, HOW THE BRAIN EVOLVED LANGUAGE 166 (1999).

[37] Chomsky relies heavily on the generativity of language to discern which from among the many patterns in human language are genuinely structural features, and which are merely accidental homologues. *See, e.g.*, CHOMSKY, REFLECTIONS ON LANGUAGE *supra* note 7, at 29; *see also* PINKER, *supra* note 30, at 32-45.

[38] *See, e.g.*, John Tooby & Leda Cosmides, *The Psychological Foundations of Culture*, *in* THE ADAPTED MIND 88-93 (Jerome H. Barkow et al. eds,, 1992); GEORGE P. FLETCHER, BASIC CONCEPTS OF CRIMINAL LAW 5 (1998) (defending the view that there is a "universal grammar" to the criminal law); Jim Chen, *Law as a Species of Language Acquisition*, 73 WASH. U. L.Q. 1263, 1279 (1995) (arguing that the "existence of universal grammar reinforces the discovery of universals in other language-based disciplines" including law); JOHN FINNIS, NATURAL LAW AND NATURAL RIGHTS 83-84 (1980); Martha Nussbaum, *Non-Relative Virtues: An Aristotelian Approach*, *in* THE QUALITY OF LIFE (Martha Nussbaum & Amartya Sen eds., 1993); A.J.M. MILNE, HUMAN RIGHTS AND HUMAN DIVERSITY 4-7 (1986); John O. McGinnis, *The Human Constitution and Constitutive Law: A Prolegomenon*, 8 J. CONTEMP. LEGAL ISSUES 211-12, 230-39 (1997) (suggesting that work in the social sciences involving evolutionary theories of human nature should begin to inform more legal scholarship and attempting to identify the deep structure of the Constitution). None of these accounts identifies as structural the precise features developed in this Article.

[39] *See* DONALD E. BROWN, HUMAN UNIVERSALS (1991).

against violence, rape and murder), punishment, conflict (which is deplored), the seeking of redress for wrongs, mediation, in-group/out-group conflicts, property, envy, and a sense of *right* and *wrong*.[40] Within the legal literature, people like George Fletcher have similarly argued that specific areas of the law, like the criminal law, exhibit a number of identifiable features in all human cultures.[41] Work like this is suggestive, though, for reasons already discussed, any attempt to distinguish between common or universal features of our moral and legal practices and those that are structural in a more robust sense, akin to Chomsky's, must draw upon additional considerations.

The second class of considerations relates to developmental psychology. There is now a well-established body of evidence that our capacities to engage in moral reasoning develop in some invariant and universal ways. These findings began with the work of Jean Piaget and Lawrence Kohlberg, the latter of whom has proposed that children, cross-culturally, progress through six identifiable stages of development in their moral reasoning, and do so in an invariant manner.[42] Experiments conducted by Kohlberg and his followers suggest by age ten, children will distinguish "right" from "wrong" but are typically motivated primarily by avoidance of punishment (Stage 1).[43] After attaining this stage, children begin to become motivated by desires for reward or benefit, and begin to develop the ability to reason instrumentally about exchanges to meet these interests (Stage 2).[44] It is then typically in early adolescence that individuals begin to develop a more robust sense of obligation.[45] This typically begins as a set of judgments based on role obligations and stereotypical conceptions of the good person and how good persons would be motivated (Stage 3).[46] Sometimes individuals develop further to make a distinctive class of judgments about obligation based on respect for uniformities, rules, law, and the authority legitimate in a social system (Stage 4).[47]

Kohlberg initially proposed two further stages,[48] but a number of

---

[40] *See id.* at 130-41; *see also* PINKER, *supra* note 30, at 413-15 (discussing Brown's work).

[41] FLETCHER, *supra* note 38. Fletcher describes this "universal grammar" as a hidden unity that underlies all criminal justice systems despite their many surface variations. For a useful discussion of how Fletcher's use of the term "universal grammar" differs from Chomsky's in relying primarily on evidence of universality, see Stuart P. Green, *The Universal Grammar of the Criminal Law*, 98 MICH. L. REV. 2104, 2108-113 (2000).

[42] 2 LAWRENCE KOHLBERG, ESSAYS ON MORAL DEVELOPMENT (1984); Lawrence Kohlberg, *From Is to Ought*, *in* COGNITIVE DEVELOPMENT AND EPISTEMOLOGY 151, 176 (Theodore Mischel ed., 1971).

[43] 2 KOHLBERG, *supra* note 42, at 52 tbl.1.6, 624-26; Anne Colby & Lawrence Kohlberg, *Invariant Sequence and Internal Consistency in Moral Judgment Stages*, *in* MORALITY, MORAL BEHAVIOR AND MORAL DEVELOPMENT 42, (William M. Kurtines & Jacob L. Gewirtz, eds. 1984)

[44] 2 KOHLBERG, *supra* note 42, at 626.

[45] *Id.* at 626, 628-29, 649.

[46] *Id.* at 628-30.

[47] *Id.* at 631-33.

[48] In Kohlberg's proposed Stage 5, individuals consider principles of fairness and equality, and reason about the right taking into account of social contracts and broad utilitarian principles. *Id.* at 634-36. At Stage 6, individuals purportedly make decisions that they derive in part from abstract universal moral principles. *Id.* at 636-39. The central tenet that sophisticated reasoning is a prerequisite for a moral maturity has, however, been challenged and is—in my view—contradicted by the weight of the evidence.

subsequent studies suggest that only developments to stage 3 are genuinely cross-cultural.[49] In smaller, rural social settings, stage 3 reasoning is the typical end-point in the development of our moral capacities, whereas developments to stage 4 reasoning arise rather robustly as people begin to live in larger, cosmopolitan settings with higher population densities and less consensus over the content of morality.[50] These facts are suggestive of capacities that may develop in distinctive ways in different social circumstances to allow us to respond to the distinctive social problems that we happen to face.[51]

Our understanding of how our moral and legal psychologies develop is still in flux and is likely to undergo a number of refinements in the coming decades.

Mary Louise Arnold has, for example, reviewed a number of studies of subjects who share highly developed commitments to moral causes. She found that "some may have perceived their commitments primarily in terms of social justice . . . and . . . may have sought to remedy them in a highly analytical, reasoned manner" while "[o]thers . . . may have perceived and responded to similar (or different) social ills with a more purely empathic or altruistic sensitivity toward human suffering." Mary Louise Arnold, *Stage, Sequence, and Sequels: Changing Conceptions of Morality Post-Kohlberg*, 12 EDUC. PSYCHOL. REV. 365, 377 (2000). "In both sets of cases, a strong moral self or personality is clearly evident, but the role of complex moral reasoning is likely far more relevant in the former cases than in the latter. Moreover, both instances present valid conceptions of moral maturity and should be equally represented and valued in moral development theory." *Id.*

In my view, Stage 5 and 6 reasoning are therefore best viewed as descriptions of culturally local attempts to reflect on stage 4 reasoning and discover its underlying principles. *See, e.g.*, 2 KOHLBERG, *supra* note 42, at 57 ("It is possible to view Stages 4, 5 and 6 as alternative types of mature response rather than as a sequence."); *id.* at 636 (defining Stage 6 by reference to Rawls's political philosophy); James Rest et al., *A Neo-Kohlbergian Approach*, 11 EDUC. PSYCHOL. REV. 291, 302-03 (1999) (noting that Kohlberg used Rawls's particular political philosophy to characterize stage 6, collecting criticisms of this linkage and of Kohlberg's interpretations of Rawls, and suggesting a broader characterization of stage 5 and 6 reasoning that would encompass a broad range of philosophical reflections on morality). Criticisms like these apply only to stages 5 and 6 in Kohlberg's taxonomy. Criticisms of this kind may, however, help explain recent moves to recharacterize Kohlberg's stages as "schema" rather than "invariant steps." *See, e.g.*, *id.* at 291, 311-319.

[49] *See, e.g.*, Carolyn P. Edwards, *Societal Complexity and Moral Development*, 3 ETHOS 505, 511 (1975); ANTHONY J. CORTESE, ETHNIC ETHICS 109 (1990).

[50] Kohlberg, *supra* note 42, at 178; Edwards, *supra* note 49, at 509-10; CORTESE, *supra* note 49, at 109 ("Stages 4/5 and 5 were absent in every traditional tribal or village folk society whether Western or non-Western. There are significant differences in moral judgment between folk and urban societies, not between Western and non-Western ones.") (collecting citations); John Snarey, *Cross-Cultural Universality of Social-Moral Development*, 97 PSYCHOL. BULL. 202, 217-18 (1985).

[51] It should be noted Piaget, Kohlberg and many of their successors did not interpret the seemingly invariant sequences of our moral development as suggestive of adaptations for a sense of obligation. *See* ELLIOT TURIEL, THE CULTURE OF MORALITY 97-98 (moral learning as the aptitude to construct more and more complex moral judgments). The data would, however, be well explained from an adaptationist perspective. The data about stages 3 and 4 are suggestive, for example, of a capacity to identify and respond to obligations that responds flexibly the needs of the social conditions that one finds oneself in. A shared sense of moral obligation may be well-suited to smaller social group interactions, whereas a sense of legal obligation, based on a sense of political authority, may be needed to resolve social contract problems arising in larger groups, which do not share a highly coordinated moral sense. If so, then we would expect legal systems to arise robustly as we move from smaller hunter-gatherer living to larger-scale settled agricultural living, with higher population densities, and this is in fact what the ethnographic record reveals. *See supra* note 49 and accompanying text.

Importantly, however, the most compelling recent developments challenge the above description only at the level of detail and ultimately reinforce the basic proposition that our moral psychologies develop in stable and predictable ways. Elliot Turiel and some of his followers have, for example, performed a number of recent and important experiments suggesting that children at a very young age distinguish between wrongs that they perceive to depend on external authority and wrongs that they perceive to be authority independent.[52] By the age of five—which is much younger than Kohlberg's theory would predict—most children will think that some wrongs—typically, those that involve harm to others—are wrong regardless of whether any authority figure such as a parent or religious entity is thought to say they are wrong.[53] These same children will, however, say that other wrongs—namely, those that do not involve obvious harms to others, such as speaking out of turn in class—are only wrong if certain authorities (such as a classroom teacher) have said so.[54] Turiel calls this the "moral/conventional" distinction,[55] and these attitudes are likely important to the distinctions we later draw between moral and legal obligations. The fact that these psychological phenomena arise robustly at specific stages in childhood development nevertheless reinforces the idea that our capacities to engage in moral and legal thought and practice have a particular developmental etiology.[56] What further research should help clarify is how precisely this capacity develops and how best to describe our capacities at various stages.[57]

---

[52] TURIEL, *supra* note 51, at 108-112 (collecting citations). One of Turiel's studies asked children aged five to eleven whether it was all right for a preschool to allow (i) hitting and (ii) the removal of clothing on warm days. The majority of the children interviewed distinguished between the two cases, and thought that only the second permission was all right. One child explained that hitting was not allowed because it "hurts other people, [and] hurting is not good," but said that a rule allowing children to remove their clothes on hot days was acceptable "because that is the rule . . . . If that's what the boss wants to do, he can do that." *Id.* at 108-09. Similar studies were conducted by Larry and Maria Nucci, who observed children responding to social breaches in the classroom and on the playground. Among their findings was the following: children typically explain their judgments about conventional wrongs with rule statements ("you're not supposed to . . .") and respond to violations with commands or ridicule, whereas children typically respond to perceived moral transgressions with retaliation and explain their judgments with injury or loss statements ("that hurt"), statements relating to perceived injustice, or statements about how it would feel to be in another's shoes. *See* Larry P. Nucci & Maria Santiago Nucci, *Children's Social Interactions in the Context of Moral and Conventional Transgressions*, 53 CHILD DEV. 403-12 (1982); Larry P. Nucci & Maria Santiago Nucci, *Children's Responses To Moral and Conventional Transgression in Free-Play Settings*, 53 CHILD DEV. 1337-1342 (1982).

[53] TURIEL, *supra* note 51, at 108-12.

[54] *Id.*

[55] *Id.* at 111-12

[56] *See also* JESSE J. PRINZ, THE EMOTIONAL CONSTRUCTION OF MORALS 14-15 (forthcoming 2005) *selections available at* http://instruct1.cit.cornell.edu/courses/phi663/PrinzECM-Excerpt.pdf (discussing these and other recent developments in the psychological literature suggesting that our moral and legal psychologies develop in predictable ways during childhood).

[57] Carol Gilligan has produced another important criticism of Kohlberg's work. Observing that some early studies suggested that men attained Stage 4 reasoning more often than women, Gilligan has argued that it would reflect a sex or gender bias to view Kohlberg's six proposed stages of moral development in terms of increasing levels of moral maturity. Gilligan has proposed that women instead exhibit a distinctive form of moral reasoning—which she characterizes as "care oriented" rather than "justice oriented"—with a distinctive etiology. *See, e.g.*, CAROL GILLIGAN, IN A DIFFERENT VOICE 18-23, 100-05 (1982).

The third category of considerations is other empirical considerations. Here too suggestive work has been done, primarily in the studies of psychopathology. These studies suggest that psychopaths often have perfectly functioning capacities of many kinds, and that what they lack is a discrete bundle of interrelated psychological phenomena that should be familiar from our moral and legal practices.[58] For example, they are (i) less capable of feeling certain characteristic moral emotions like remorse, shame and guilt;[59] (ii) less capable of empathy and role-taking;[60] and (iii) less capable of perceiving what the rest of us take to be the distinctive authority or compelling nature of law and morality.[61]

With regard to this last point, psychopaths do seem capable—at least to some extent—of learning what people call "right" and "wrong" (or "required by law" and "against the law") in their society, and of learning what reactions typically follow from conduct that is in this sense prohibited.[62] They will also often use the special normative terminology that most of us use to make moral and legal judgments to refer to such conduct. When psychopaths make their judgments, they do not, however, appear intrinsically moved by them in the way that most of us are: they are neither moved to respond directly to morality or law, though they can be moved by thoughts of potential sanctions, nor do they understand these critical reactions to be warranted by compelling reasons.[63] These further implications appear to be part of what we mean when we

---

Subsequent experiments have, however, called a number of Gilligan's factual assumptions into question. *See, e.g.*, L.J. Walker, *Sex Differences in the Development of Moral Reasoning*, 55 CHILD DEV. 677-91 (1984) (presenting meta-analysis of empirical literature suggesting that sex differences in Kohlberg's stage levels disappeared once education and occupation were controlled); Eva E. A. Skoe et al., *The Influence of Sex and Gender-Role Identity on Moral Cognition and Prosocial Personality Traits*, 46 SEX ROLES 295, 304-07 (2002); CORTESE, *supra* note 49, at 99-101. Some of the research suggests that, in fact, both men and women rate care dilemmas as more important that justice dilemmas, and that women are, all other things equal, better at both Stage 3 and Stage 4 reasoning. *See, e.g.*, Skoeet al., *supra*, at 301-02, tbl.I, 307. To the extent that fewer women exhibit Stage 4 reasoning, it is tempting to conclude that this reflects a familiar cultural failure: many cultures limit the social circumstances in which women participate, such that Stage 3 reasoning may naturally become predominant for those persons. This may occur in any culture that channels women into "caring" activities and a focus on the family, and not as frequently into the political sphere where negotiating relationships with outsiders is required. *See, e.g., id.* at 305.

[58] *See generally* PRINZ, *supra* note 56, at 16-20, TURIEL, *supra* note 56, at 16-20 (describing the psychology of psychopaths); HERVEY CLECKLEY, THE MASK OF SANITY (4th ed. 1964) (same).

[59] *See, e.g.*, ROBERT D. HARE, WITHOUT CONSCIENCE – THE DISTURBING WORLD OF THE PSYCHOPATHS AMONG US 33-57 (1999) ("Psychopaths show a stunning lack of concern for the devastating effects their actions have on others. Often they are completely forthright about the matter, calmly stating that they have no sense of guilt, are not sorry for the pain they have caused, and that there is no reason for them to be concerned."); PRINZ, *supra* note 56, at 16.

[60] *See, e.g.*, HARE, *supra* note 59 ("[Psychopaths] seen unable to 'get into the skin' or to 'walk in the shoes' of others, except in a purely intellectual sense."); JAMES Q. WILSON, THE MORAL SENSE 107-08 (1993); PRINZ, *supra* note 52, at 16.

[61] *See, e.g.*, Jeffrie G. Murphy, *Moral Death: A Kantian Essay on Psychopathy*, 82 ETHICS 284, 286 (1972) (noting that psychopaths are not moved by morality or law); PRINZ, *supra* note 56, at 16-17

[62] Murphy, *supra* note 61, at 286; PRINZ, *supra* note 56, at 17-20.

[63] Murphy, *supra* note 61, at 286; PRINZ, *supra* note 56, at 17-20.

sincerely take moral and legal obligations to have their distinctive kinds of authority. And these differences between psychopaths and ordinary persons would be well-explained if we had a distinctive adaptive capacity to identify and respond appropriately to obligations, which can be selectively disrupted in some individuals.

It is, however, at this critical point that any proposed analogy between language, on the one hand, and law and morality, on the other, breaks down. The fourth set of considerations that support Chomsky's views on language relate to its generativity, and the distinctive function that human language serves in allowing us to express indefinitely many thoughts by means of a discrete combinatorial system.[64] This has led many, including John Rawls most prominently, to suggest that an analogous theory of morality and our sense of justice should contain a "generative grammar," in approximately Chomsky's sense.[65] But a careful look at our moral and legal judgments suggests that they do not reflect any distinctive problem of generativity that could be properly analogized to that of language. We do make all kinds of moral and legal judgments—saying, for example, that some things are "wrong" or are "against the law"—and we can embed these judgments into more complex syntactical structures, thus revealing that our moral and legal thought is as richly flexible as our thought in any other domain. This dimension of flexibility is, however, wholly derivative of that of language, and tells us nothing specific about morality or law. Moreover, the more basic judgments upon which these general grammatical rules operate appear to be ascribing properties like "rightness" or "legality" to various actions. If we are to explain how children can learn to apply properties like these to indefinitely many situations based only on a finite number of learning stimuli, we must—as Rawls has also noted—presume some psychological capacities to resolve "what linguists and cognitive scientists call the 'projection problem'—roughly, "the ability of normal persons to make systematic and stable moral judgments about an indefinite number of cases of 'first impression,' *i.e.*, novel fact patterns falling outside their previous experience."[66] This same problem arises whenever we learn to apply the concept for a given property, however, and does not yet raise the need of a discrete combinatorial system or generative grammar. To see this, notice that this class of problems would arise for any property even if our linguistic capacities involved only simple thoughts ascribing properties to actions or events, and did not yet give us the capacity to embed these thoughts into more complex syntactic structures. Hence, the capacities we use to resolve the projection problems inherent in moral and legal thought do not appear to involve resolutions of any particular problems of generativity.

There is no point forcing an analogy beyond its power to illuminate. What would be needed to render plausible the claim that morality and law share a deep

---

[64] *See supra* notes 30-32 & accompanying text.

[65] *See* JOHN RAWLS, A THEORY OF JUSTICE 47 (1971); John Mikhail, *Law, Science, and Morality: A Review of Richard Posner's* The Problematics of Moral and Legal Theory, 54 STAN. L. REV. 1057, 1091 n.204 (2002). Chomsky has himself suggested that his work in linguistics might provide a "suggestive model for inquiry into other domains of human competence and action that are not quite so amenable to direct investigation." CHOMSKY, REFLECTIONS ON LANGUAGE, *supra* note 7, at 5.

[66] Mikhail, *supra* note 65, at 1090 n.202 (2002) (comparing relevant passages from Rawls and Chomsky for this proposition).

structure is not a defense of the (ultimately implausible) claim that they function as discrete combinatorial systems but rather the identification of a genuine and specific function that our moral and legal psychologies share, along with a compelling argument that clarifies how serving this function would require certain psychological conditions that in fact pervade our moral and legal practices in the way that would be predicted by such a function. The relevant function—or so this Article will argue—is that morality and law allow us to resolve social contract problems of a variety of forms, and to do so in a flexible manner. The main sections of this Article will aim to establish this point, and to trace out its implications for our understanding of which precise features of our moral and legal practices are genuinely "structural" in the sense that they are universal, recurrent and pervasive conditions for our moral psychologies to serve their natural functions in a stable manner.

There is a long and familiar tradition of trying to account for moral or political right in social contractarian terms.[67] This Article will not, however, depend on any such normative considerations to argue for the particular natural function that it attributes to our moral and legal psychologies. The main argument will instead simply begin by asking what kinds of capacities we would need to resolve social contract problems. The Article will then proceed in stages by clarifying various features that we should expect of any such capacities and then testing these predictions against the available evidence from a wide range of sources, including: moral and legal philosophy, anthropology, psychology, primatology (and other animals studies where relevant), ordinary language and evolutionary theory. Each layer of the argument will provide added support for the emerging conclusion that the proposed function is in fact the natural function of the attitudes that breathe life into our moral and legal practices, and each layer will illuminate further features of our moral and legal psychologies that should thus be deemed "structural" in the relevant sense.

Before turning to this project, one final clarification is in order to prevent an important misunderstanding. This Article will sometimes draw on arguments rooted in evolutionary psychology and evolutionary game theory to clarify structural features of our moral and legal psychologies. The use of such resources to explain features of our

---

[67] Social contractarians account for moral or political right in terms of "principles that are, or would be, the object of a suitable agreement between equals." Stephen Darwall, *Introduction*, *in* CONTRACTARIANISM/CONTRACTUALISM 1 (Stephen Darwall ed., 2003). Classical social contract theorists include Thomas Hobbes, John Locke, and Jean-Jacques Rousseau. *See, e.g.*, THOMAS HOBBES, LEVIATHAN (W.W. Norton & Co. 1997) (1651); JOHN LOCKE, TWO TREATISES OF GOVERNMENT (Peter Laslett ed., 1988) (1690); JEAN-JACQUES ROUSSEAU, THE SOCIAL CONTRACT (Great Books Foundation 1948) (1762). Important contemporary versions of the view have been defended by Stephen Darwall, David Gauthier, John Rawls, and T.M. Scanlon. *See, e.g.*, STEPHEN DARWALL, MORALITY AND THE SECOND-PERSONAL STANDPOINT (forthcoming, on file with author); DAVID GAUTHIER, MORALS BY AGREEMENT (1986); JOHN RAWLS, A THEORY OF JUSTICE (1971)*;* JOHN RAWLS, POLITICAL LIBERALISM (1993); SCANLON, *supra* note 11. The particular views of these theorists often differ along a number of important dimensions, including: (i) whether they use a contractarian apparatus to account for moral or political/legal authority, or both; (ii) whether they make reference to actual or hypothetical agreements; (iii) how they picture the relevant original bargaining situation (sometimes called the "state of nature" or "original position") and the equality inherent in it; and (iv) how they picture the psychologies of the relevant bargainers. A more recent distinction in the literature is between "'*contractarianism*, where the parties' equality is merely *de facto* and their choice of principles rationally self-interested, and *contractualism*, which proceeds from an ideal of *reasonable* reciprocity or fairness between *moral* equals." Darwall, *Introduction*, *supra*.

social psychologies might raise familiar concerns given the number of questionable uses to which evolutionary theory has been put to purportedly justify substantive normative proposals. As examples, one need only think of (i) the eugenics movement, which drew heavily on evolutionary conceptions of fitness purportedly to justify the intentional and systematic eradication of some persons from our ongoing gene pool;[68] (ii) Herbert Spencer's claim that laissez-faire economics could be justified in terms of evolutionary concepts like survival of the fittest;[69] or, indeed, (iii) the use of evolutionary theory to purportedly justify everything "from the extermination of ethnic groups and the forced sterilization of the poor to restrictive immigration laws and legally institutionalized sex and race discrimination."[70] A distinct but related concern might arise if evolutionary theory were being used being used to try to cabinet what is morally or legally possible for us.[71]

None of the claims in this Article should raise these familiar concerns. Although this Article will identify a number of structural features of our moral and legal psychologies, these claims will be perfectly consistent with the broadest range of moral and legal views. It will, for example, be logically consistent with everything that is said here for a given culture or an individual to believe that any particular thing is morally required or prohibited, or, even, that everything is morally permissible. The same full range of logical possibilities will arise in terms of the possible contents of the law. Moreover, nothing in the explanatory claims will by themselves foreclose any particular normative positions on what moral or legal positions are legitimate, correct, worth cultivating or otherwise conducive to human welfare. In my view, the fact that we can identify and respond appropriately to obligations characterizes a distinctive dimension of human freedom and human possibility. Hence, the claims in this article should be viewed not as placing limitations on human freedom but as clarifying what a particular dimension of human freedom amounts to.[72]

---

[68] *See generally, e.g.*, MARK H. HALLER, EUGENICS: HEREDITARIAN ATTITUDES IN AMERICAN THOUGHT 3-4 (1963) ("[E]ugenics was the . . . offspring of Darwinian evolution, a natural and doubtless inevitable outgrowth of currents of thought that developed from the publication . . . of Charles Darwin's *The Origin of Species*."); DANIEL J. KEVLES, IN THE NAME OF EUGENICS 12, 180-83 (rev. ed. 1995) (chronicling the use of evolutionary thinking to support various eugenicist programs).

[69] *See generally, e.g.*, PETER SINGER, A DARWINIAN LEFT: POLITICAL, EVOLUTION AND COOPERATION 11 (2000) ("Herbert Spencer, who was more than willing to draw ethical implications from evolution, provided the defenders of laissez-faire capitalism with intellectual foundations that they used to oppose state interference with market forces.").

[70] Tooby & Cosmides, *supra* note 38, at 34-35.

[71] These concerns would naturally dovetail if evolutionary theory were being used to suggest that certain facts about us that would otherwise be normatively objectionable are in some sense necessary and unavoidable and therefore justifiable. Factors like these make me very sympathetic to Phil Kitcher's way of describing what he calls "the history of brave, but disastrous, ventures into evolutionary ethics." Phil Kitcher, *Psychological Altruism, Evolutionary Origins, and Moral Rules*, 89 PHIL. STUD. 283, 283 (1998).

[72] Indeed, there are some views of freedom—those in the broadly Kantian tradition—that would deem human freedom, or autonomy, to arise only once we have the capacity to step back from our antecedent desires and interests and respond to reasons that are categorical in the way that moral and legal obligations purportedly are. For the classic statement of this view, see IMMANUEL KANT, GROUNDWORK FOR THE

On the other hand, a clear understanding of how these capacities function is likely to help clarify what would be involved in normative proposals that would engage these capacities. Here, an analogy with language is again helpful. The deep structure of language is—on Chomsky's view—part of what gives us the capacity to understand and express the rich variety of thoughts that we see in all languages, and to learn and use languages so naturally. One might try to construct a system of communication for the same purpose that does not employ these natural capacities, but the evidence suggests that such systems of communication require excessive conscious processing and thought, are difficult to learn and unstable in human memory, and do not have nearly the richness and flexibility of expression of languages that directly employ our native linguistic capacities.[73] While our capacities to identify and respond appropriately to obligations may give us the freedom to understand obligations with any particular content, a similar point may apply to our moral and legal practices. The attitudes that allow us to respond to moral and legal obligations are—for reasons to be explained below—a bundle of psychological phenomena, which tend to come together as part of a distinctive syndrome. Hence, while normative assessments about the appropriate content of morality and law may be useful, and while it may be useful to discuss the appropriate roles of morality and law in our lives, it may be very difficult for us to sustain normative proposals that would require engagement of parts of our moral or legal psychologies along with an abandonment of the rest. As with the analogue in language, we may be able to respond to such normative proposals only haltingly, with great difficulty, or in an unstable manner. If so, then these are facts that we should understand about ourselves. An understanding of these facts will likely bear on what normative questions about morality and law are genuine and live.

B. A STARTING DILEMMA: OUR CAPACITIES TO RESOLVE SOCIAL CONTRACT PROBLEMS

METAPHYSICS OF MORALS (Cambridge Univ. Press 1997) (1785); *see also* CHRISTINE KORSGAARD, THE SOURCES OF NORMATIVITY (1996).

[73] CHOMSKY, REFLECTIONS ON LANGUAGE *supra* note 7, at 29 ("If we were to construct a language violating [universal grammar], we would find that it could not be learned [as a human language]. That is, it would not be learnable under normal conditions of access and exposure to data. Possibly it could be learned by application of other faculties of mind; [human languages meeting the requirements of universal grammar do] not exhaust the capacities of the human mind. This invented language might be learned as a puzzle, or its grammar might be discovered by scientific inquiry over the course of generations . . . . But discovery of the grammar of this language would not be comparable to language learning . . . ."). One might get a sense of the limitations inherent in communication without deep grammar by observing so-called "pidgin" languages. These languages typically arise whenever two groups with different native tongues are forced to interact and work with one another without a shared language. Pidgins are primitive attempts to communicate, typically by using strings of words from one or another native language. Pidgins have little in the way of grammar to facilitate the full range of expression to which we are accustomed, however, and pidgins can only be used for relatively simple tasks. These languages should thus be contrasted with those that second generation speakers develop when exposed to pidgins during their crucial language learning windows. These children typically develop fully grammatical and expressive languages called "creoles," which do allow for the full range of expression of natural language. PINKER *supra* note 30, at 33-35.

This section begins the substantive arguments of the Article with a thought experiment. Imagine that we had all of the ordinary psychological capacities that we now have but no adaptive capacity that functioned to allow us to resolve social contract problems. We would presumably still have all kinds of ordinary cognitive capacities, and could therefore form all kinds of beliefs about the world. We would also presumably still have a broad array of motives, including most of what we typically think of as "desires." It is, in fact, an open question—at least at this stage—whether we would be lacking anything at all.

Given these starting assumptions, one might also develop a plausible account of practical reasoning—or of how we reason about what to do—that is wholly consistent with a broadly naturalistic worldview. Following Stephen Darwall, one might define an "agent's reasons for action" as those considerations that both (i) explain an agent's action and (ii) explain it from a perspective internal to the agent or as an expression of the agent's conception of what seemed to speak in favor of the alternative at the time.[74] Reasons like this are thus either causes or intimately bound up with causes, which motivate actions.[75] It can be deeply puzzling how to fit reasons for action into a naturalistic worldview, but, as a number of people have observed, desiring something typically involves a tendency to see it as desirable or good, or as providing one with a reason to pursue it.[76] When coupled with relevant beliefs, desires of this kind might therefore motivate an action and allow us to explain it as an expression of the agent's conception of what seemed to speak in favor of it.

To take a stock example, a person might desire an apple that she sees on a tree, and, in desiring it, see it as desirable or good. She might also believe that she would need help to get the apple. Practical reflection might therefore lead the person to decide to seek help to get the apple, which decision might translate into the help seeking action. If so, then the person's action could be explained in an unproblematic way by the belief and desire under discussion. Of course, the person herself would be taking the *object* of her belief (namely, the fact that she needs help) and the seeming *desirability* of the apple as her reasons for action.[77] But these perceived facts would have the right kind of motivational force needed to explain her actions because these thoughts would themselves be in part expressive of the belief and desire under discussion. It is, in fact, common to think that human actions typically arise from a combination of beliefs and desires in roughly this manner.[78] Support for this view derives from the plausibility of

---

[74] *See, e.g.*, STEPHEN DARWALL, IMPARTIAL REASON 32-34 (1983).

[75] For the classic argument that reasons are also causes, see Donald Davidson, *Actions, Reasons and Causes*, *in* ESSAYS ON ACTIONS AND EVENTS 5-19 (1980).

[76] *See, e.g.*, Warren Quinn, *Putting Rationality in Its Place*, *reprinted in* MORALITY AND ACTION 246-47 (1993); SCANLON, *supra* note 11, at 38.

[77] This last point is sometime made by saying that desires are in the "background," rather than the foreground, of practical deliberation. *See, e.g.*, Philip Pettit & Michael Smith, *Backgrounding Desire*, 99 PHIL. REV. 565-92 (1990).

[78] For the classic description of this view in the philosophical literature, see Davidson, *supra* note 75, at 3-19; *see also* MICHAEL E. BRATMAN, INTENTION, PLANS, AND PRACTICAL REASON 5-7 (1999) (describing intuitions that appear to warrant adoption of a belief-desire account of practical reason, along with a number of its proponents). This is also the most common view of rational choice that most economists

the psychological and causal stories it proposes, and from its consistency with a broadly naturalistic worldview.

There is, however, room for reasonable disagreement at this point as to whether there are any objective facts about desirability or our personal good that explain our corresponding beliefs or if these beliefs are merely projections of our desires. Presumably, only entities that are part of the natural world can have causal powers. Hence, in order to elaborate an objectivist view, one would need to specify natural facts that plausibly constitute a person's objective good (or what is objectively desirable for her) along with a plausible psychological account of how knowledge of these facts might motivate action. One would also need to provide a plausible naturalistic account of how we might have epistemic access to these particular facts. Peter Railton has provided the most promising account of the relevant type.[79] He has articulated a plausible naturalistic account of a person's objective interests,[80] and has collected a number of broad theoretical grounds for believing that we have what he calls a "wants/interests mechanism," or a psychological mechanism that reorients our desires, however imperfectly, and "permits individuals to achieve self-conscious and un-self-conscious learning about their interests through experience."[81] If this hypothesis is correct, then facts about what is objectively in our interests may explain our actions in part by engaging desires that make us see certain things as desirable or good.

Others—like David Gauthier—have expressed skepticism about the notion of an objective interest.[82] For skeptics, our beliefs that certain things are good or desirable are often better thought of merely as projections of our desires, with nothing further to ground them.[83] Still, the idea that these desires give us reasons for action is often thought unproblematic, and, in fact, an agent's good or utility is often defined, more subjectively, in terms of the satisfaction of such desires.[84] Most of the definitions of utility that are common in the economics literature fall into this category, as they define an agent's utility or good in terms of the maximization of things like considered

---

presume. *See generally* Albert Weale, *Homoeconomicus, Homosociologicus*, *in* HEAP ET AL., THE THEORY OF CHOICE 62-72 (1997) (describing numerous applications of this model in the economics literature). There is a good amount of evidence establishing that we deviate from the description of rational choice that economists have assumed in numerous and systematic ways. *See generally* Robert Sugden, *How People Choose*, *in* HEAP ET AL., *supra*, at 36-50 (describing a number of these developments). The idea that desires are the only things that can provide us with reasons for action is also hotly contested. For a good critique of this view, see Darwall, *supra* note 21, at 129-53.

[79] Peter Railton, *Moral Realism*, 95 PHIL. REV. 163 (1986), *reprinted in* PETER RAILTON, FACTS, VALUES, AND NORMS 9-17 (2003).

[80] *Id.* at 10-13.

[81] *Id.* at 14, 15-17.

[82] David Gauthier, *Why Contractarianism?*, *in* CONTRACTARIANISM AND RATIONAL CHOICE 15 (Peter Vallentyne ed., 1991) *reprinted in* CONTRACTARIANISM/CONTRACTUALISM, *supra* note 67, at 91, 94.

[83] Gauthier, for example, goes to great lengths to argue that the reasons that issue from practical deliberation from this kind of source do not need a foundation. *Id.* at 94-96.

[84] *See, e.g.*, *id.* at 94.

preferences.[85] Theorists in this vein often assume that practical rationality requires us to maximize our individual utility in this sense.[86]

Whether in objectivist or subjectivist terms, the psychological assumptions under discussion can thus be used to define a metric for individual desirability or interest, either rooted in or related to a person's considered beliefs and desires. Importantly, however, nothing has been said that explains how we might identify and respond appropriately to obligations, as opposed to these ordinary reasons for action.[87] It is, moreover, common to think that obligations prove a much more problematic notion. As indicated earlier, I believe an answer to this question will arise, indirectly, from a clarification of the capacities we would need to resolve social contract problems as defined by the metric for individual desirability under discussion here. Social contract problems have the underlying game-theoretic structure of an *n*-person prisoners' dilemma—a decision situation that many take to be of central importance to moral, legal, political and social philosophy.[88] This section will thus proceed by asking what capacities we would need to resolve these problems, and will focus the analysis by drawing attention to a number of familiar problems that would arise if, as is sometimes assumed, we must resolve them starting from a blank slate, armed only with separable beliefs and desires to decide what to do. This picture of human psychology and interaction is one that will ultimately be rejected, but some of the problems it raises will help motivate the alternative this Article will propose.

Let us turn, then, to social contract problems. Social contract problems have a well-known structure, which would—as it turns out—make it extraordinarily difficult to understand how we could ever have the internal capacity to resolve them from this kind of starting point. Social contract problems arise whenever we could all do better by agreeing to be bound by some standard of action if that were the price of having all (or a significant majority of) others be similarly bound. The resolution of such problems thus indisputably requires a capacity on the part of each to be motivated to act in accordance with the relevant shared standards for action on the condition that all (or a

---

[85] *See, e.g.*, RICHARD POSNER, ECONOMIC ANALYSIS OF LAW 3-4 (6th ed. 2003); LOUIS KAPLOW & STEVEN SHAVELL, FAIRNESS VERSUS WELFARE 409-31 (2002).

[86] *See, e.g.*, ROBERT COOTER & THOMAS ULEN, LAW & ECONOMICS 10-11 (2000); Shaun Hargreaves Heap, *Rationality*, *in* HEAP ET AL., *supra* note 78, at 3, 4.

[87] The distinction between obligations and reasons has not always been sufficiently appreciated in this context, and it is sometimes assumed that a general capacity to respond to reasons will give us the capacity to respond to obligations, because obligations present us with a particular kind of reason. As discussed more fully below, this assumption obscures important details about how our moral psychologies function.

[88] There are some who have argued that other game-theoretic situations better capture the nub of certain social contract problems. Jean Hampton's analysis of Hobbes, for example, suggests that the Hobbesian contract may present more of an assurance problem than a prisoners' dilemma. *See* JEAN HAMPTON, HOBBES AND THE SOCIAL CONTRACT TRADITION 197-207 (1986). Robert Sugden has similarly argued that property conventions are best modeled as hawk-dove games. *See* Robert Sugden, *Normative Expectations: The Simultaneous Evolution of Institutions and Norms*, *in* ECONOMICS, VALUES, AND ORGANIZATION 73-100 (Avner Ben-Ner & Louis Putterman eds., 1998). For reasons that will become clear below, however, starting with an account of social contract problems as presenting us with *n*-person prisoners' dilemmas will clarify how exactly these other game-theoretic problems might arise and be resolved in particular cases. These alternative proposals thus do not cast doubt on the fruitfulness of this proposed starting point.

significant majority of) others are similarly motivated. What is less obvious, but equally true, is that the sustained resolution of these problems requires, in addition, a mutually concordant system of expectations among the members of the group concerning one anothers' motivations, one that inclines the group toward a relevant cooperative (rather than defecting) equilibrium.[89] Only with such a mutually concordant system of expectations in place will the perceived conditions needed to engender the shared motives to act in accordance with the social contract persist.

Social contract problems are also a species of what has in the last several decades increasingly come to be known as *commitment problems*.[90] These are typically defined as any dynamic, strategic problem in which an individual can obtain more desirable or self-interested results by giving up certain options or by guaranteeing others—in short, by making commitments.[91] Commitment problems arise whenever a commitment is the price of some action or reciprocal commitment on the part of one or more others. This kind of problem can also arise when the threat of a commitment is needed to change the way one or more others will behave. Hence, according to Thomas Schelling, "[t]o commit is to relinquish some options, eliminate some choices, surrender some control over one's future behavior—and [to do] so with a purpose. The purpose is to influence someone else's choices."[92] There is broad agreement in the literature on commitment problems that commitments are strategies that work by changing what others believe about us.[93] In the case of social contract problems, this orthodox view

---

[89] This is a *kind* of "assurance problem," though not the standard one that is typically referred to by that name in the literature. In the standard assurance game, there are one or more joint decisions that each would prefer to participate in on condition that all others will play their respective parts in the same joint decision. *See, e.g.*, Carlisle F. Runge, *Institutions and the Free Rider: The Assurance Problem in Collective Action*, 46 J. POL. 154-55, 158, 160-61 (1984) (contrasting assurance problems with prisoners' dilemmas). Hence, none has an interest in defecting and there is no free rider problem. In the situation canvassed in the main text, by contrast, each is stipulated to have an interest in defecting along with a motivational capacity to cooperate on condition that all (or a significant majority of) others are similarly motivated. Assuming these latter motivational capacities, there is, however, still an outstanding question of what might assure each that he or she is in the kind of circumstances of generalized reciprocity that would invite personal cooperation—thereby allowing for cooperation to be sustained. This problem bears important resemblances to the standard assurance problem, but the two problems are not strictly identical.

[90] *See generally* Randolph M. Nesse, *Natural Selection and the Capacity for Subjective Commitment*, *in* EVOLUTION AND THE CAPACITY FOR COMMITMENT 3-44 (Randolph M. Nesse ed., 2001) [hereinafter ECC] (identifying commitment problems and describing their growing importance in the social, economic and biological literature).

[91] *Id.* at 12-18; Jack Hirshleifer, *Game-Theoretic Interpretations of Commitment*, *in* ECC, *supra* note 90, at 78-91.

[92] Thomas C. Schelling, *Commitment: Deliberate Versus Involuntary*, *in* ECC, *supra* note 90, at 48

[93] Randolph Nesse says of Schelling, Frank and Hirshleifer—the three contributors to the theoretical section of his volume *Evolution and the Capacity for Commitment*—that "[a]ll of them emphasize . . . that commitments are interesting mainly when they are for actions that would not otherwise be expected." Randolph M. Nesse, *Core Ideas From Economics*, *in* ECC, *supra* note 90, at 45. In his contribution, Hirshleifer says that "[t]o be effective, a commitment not only must be made but conveyed." Hirshleifer, *supra* note 91, at 87. Frank's contribution canvasses the importance of emotional commitment mechanisms, which he describes as operating by means of adaptive signals that let others know what we are feeling. *See* Robert H. Frank, *Cooperation Through Emotional Commitment*, *in* ECC, *supra* note 90, at 57-64.

would entail that we solve these problems by committing ourselves, conditionally, to the terms of a social contract, and by expressing our conditional commitments to one another in order to let everyone know that we are willing partners and thereby induce everyone else who is willing to make the relevant reciprocal commitments to us.

But there is a difficulty inherent in this orthodox view. For how exactly *are* we to do or say anything that could convey this kind of information, something that could serve as evidence, in some recognizable sense of the word, for a belief that the conditions required for us to exhibit our commitments are in place? Given our initial assumptions, an explicit agreement, consisting of reciprocal promises, could not serve as a rational basis for the formation of a mutually concordant system of expectations on its own. This is because the value of a capacity to make promises is said to derive from its ability to induce others to do things that they would not otherwise do. But reciprocal promises could only justify promisors in forming the belief that reciprocal commitments have been taken on if they already expect one another to be committed to a rule of promise keeping. There is thus the prior and familiar question as to what would justify these expectations, and no explicit agreement could do this work alone on pain of regress. If explicit agreement cannot alone form the rational basis for a relevant system of mutually concordant expectations in these circumstances, and if each member's commitments are conditioned on such expectations, then how can social contract problems be resolved in the first instance?[94]

One possibility should be set aside from the start. There are sometimes circumstances in which we can give up certain options or guarantee others in a sufficiently trustworthy manner simply by altering features of our external situation, such that different alternatives become more or less possible or desirable. The use of collateral to secure performance is an example of this kind of resolution, as is Hobbes's proposal that we establish a Leviathan with sovereign power to enforce the terms of our social contracts.[95] In cases like these, the commitment is secured by changing features of our external situation, however, and, hence, without the need for any internal capacity to fulfill our commitments and/or refrain from acting on our desires. Unfortunately, security mechanisms are not always available, and some—such as the establishment of a political authority—can themselves require the resolution of a collective action problem that has precisely the same form as the social contract problem that it was originally meant to resolve.[96] The important point to recognize for the present purposes is that

---

[94] I do not mean to suggest, of course, that we cannot actually solve social contract problems by relying on explicit agreement. We do this all the time. The arguments in the main text nevertheless clarify how difficult it can be to understand how we might do this given only the psychological assumptions that are in play—a fact that should cast doubt on the assumptions themselves and help clarify how, precisely, they should be modified. A similar remark will apply to the rest of the arguments in this section.

[95] Hobbes' classic statement of this view is in *Leviathan, supra* note 67.

[96] Jean Hampton has argued that this fact renders Hobbes's account of how people in a state of nature might establish an absolute sovereign inconsistent with his particular psychological assumptions, which allow for overriding motives of self-preservation to persist in the State. Hampton rests her case in large part on a more general line of argument: although it can be rational for expected-utility maximizers to authorize a political sovereign to enforce the terms of a social contract, such authorization involves submission not only to ordinary sovereign orders but also to sovereign commands to punish others for breaches, and, more generally, to the sovereign's will, including its will not to be frustrated in any of its enforcement powers. But because such submission can require action against individual self-interest or

security mechanisms ultimately allow us to dissolve social contract problems, not solve them. Perhaps what we call "obligations" are, upon closer examination, nothing more than systems of motives supported by security mechanisms "all the way down." But if so, then we would not need an internal capacity to respond to these so-called "obligations," and the distinction between systems of "obligation," in this sense, and systems of coercion would ultimately come to naught.[97] What this shows is that if we have an internal capacity to resolve social contract problems, the capacity cannot consist in our use of any such security mechanisms, or of their known consequences, to generate the relevant systems of mutually concordant expectations.[98]

One might think that ordinary induction could instead do the trick of producing the relevant system of expectations. There is, however, a pragmatic problem with this suggestion. Even if we were all conditionally committed to a rule of promise keeping, or, more directly, to a set of norms for mutual advantage that we might contract into by relying on such a shared rule, the initial absence of the relevant system of expectations would mean that none of us would expect the conditions to obtain in which our commitments would lay claims on our individual conduct. This gap in expectation would, in turn, become a kind of self-fulfilling prophecy: with no expectations that these conditions obtain, none of us would take ourselves to be obligated to act in accordance with the rules in question, and, hence, none of us would exhibit any committed behavior.[99] But then none of the conditions needed to engender our commitments would exist, and few, if any, commitments would show up to experience.

Here is a perhaps too-quick conclusion that we might draw from these considerations, but one that will be given additional justification in the ensuing sections: In order to solve social contract problems, we apparently need a mutually concordant system of expectations that runs ahead of the evidence and is, perhaps, even in some

---

utility, no psychology allowing only for expected-utility calculations relating to whether to obey sovereign commands, including sovereign enforcement commands, could allow for the creation of an entity with such sovereign power. *See* HAMPTON, *supra* note 88, at 197-206. Jon Elster has argued, similarly, that groups of rational utility-maximizers cannot resolve collective action problems through the creation and maintenance of systems of sanctions, because—in his view—it is generally costly for agents to engage in sanctioning activities or express disapproval, and, hence, something besides utility-maximization must generally sustain a norm of sanctioning. *See* JON ELSTER, THE CEMENT OF SOCIETY 132-33 (1989).

[97] H.L.A. Hart famously drew on considerations like these to develop his particular account of law in THE CONCEPT OF LAW. More specifically, he drew heavily on the fact that our concept of "obligation," as opposed to that of being "obliged" at gunpoint, cannot easily be analyzed in terms of habits of obedience to sovereign commands backed by coercive sanctions, and must instead include reference to what he called the "internal point of view." *See* HART, *supra* note 5, at 26 61. In Hart's view, those who take themselves to be under an obligation view the relevant imperative as providing them with not only an internal guide to conduct but also a reason to criticize deviations. *See id.* at 51-61, 82-91. In addition, obligations arise only where "the general demand for conformity is insistent and the social pressure brought to bear upon those who deviate or threaten to deviate is great." *Id.* at 86. For another classic argument distinguishing obligation from coercion, see ROUSSEAU, *supra* note 67, at 49-51, 52-59.

[98] This does not mean that security mechanisms cannot provide additional support to motivations that are part of an internal capacity. It does mean, however, that the capacity must not depend solely on such security mechanisms.

[99] Peter Railton has discussed this potential problem in *Getting Started: The Problem of Regress*, *in* REASON AND VALUE: THEMES FROM THE MORAL PHILOSOPHY OF JOSEPH RAZ (2004)

ways recalcitrant to the evidence. We apparently need something closer to shared attitudes of unsecured trust than individual beliefs based on evidence of trustworthiness.[100] But this suggestion only makes the notion that we might have an internal capacity to solve social contract problems even more mysterious. For what could warrant this trust, if not evidence of trustworthiness, and what guarantees that any unsecured expectations we might have would form part of a mutually concordant system? Trust is, moreover, ordinarily sensitive to evidence of trustworthiness, and, indeed, it would seem to have to be to serve its ordinary function. So what can this proposed relation to evidence amount to if not blind and misdirected trust?

At the same time, overcoming these obstacles is critical for many of the more familiar ways in which we resolve social contract problems. Expressing reciprocal promises and engaging in many other analogous actions may in fact work quite well in ordinary life, but the arguments in this section suggest that they do so only against the backdrop of certain shared psychological facts that can seem quite mysterious from a naturalistic perspective.

C. A FIRST BLUSH RESOLUTION: OBLIGATA

A careful look at the obstacles to resolving social contract problems that the last section highlighted will begin to point to a resolution. We could solve social contract problems, without security mechanisms, if we had an internal capacity that employed a particular constellation of attitudes, each fixated on a common set of contents that purported to attribute a particular property to all actions falling under a shared standard. As a first blush functional description, the constellation—which will be called an *obligatum* for ease of reference—would have to consist, at minimum, in (i) motivations to perform the act with the purported property on the supposition that all (or a significant majority of others) are similarly motivated; (ii) suppositions that all others are similarly motivated; and (iii) suppositions that all others suppose all others to be similarly motivated, and, hence, that all others have suppositions that would, if true, show them to be in circumstances that we might recognize as ones of generalized reciprocity. Contents specifying that a particular action (or its absence) had the property in question in a specific set of circumstances would reflect the thought that the relevant action is required (or forbidden, respectively) in those same circumstances.[101] If we were to share obligata with the same contents, and if the contents were to embody resolutions to a social contract problem, then we would begin our encounters with one another with the suppositions needed to trigger our respective motivations to fulfill the terms of the relevant social contract.

This and the following sections will argue that we have obligata, the contents of which naturally tend toward mutually concordant systems of motive and

[100] For another set of considerations leading to this same conclusion, see Karen Jones, *Trust as an Affective Attitude*, 107 ETHICS 4-25 (1996). Jones presents a plausible case for the proposition that trust can directly and favorably motivate those who are counted on, thus helping to sustain cooperation. *Id.*

[101] The attitude might reflect one of permission with respect to all other actions in all other circumstances; or it might be indeterminate, thus leaving other attitudes to determine whether these other actions in other circumstances are required, prohibited, or, ultimately, permitted.

supposition, and that an understanding of these attitudes is crucial for understanding how moral and legal obligations operate. To sustain this claim, this first blush definition will have to undergo a number of important refinements. To foreshadow, the final definition will stipulate that obligata include (iv) not ordinary suppositions but ones that will be called "normative suppositions" (which involve a certain amount of unsecured trust); and (v) certain secondary stabilizing attitudes. Later sections will argue that obligata provide us with reasons that we perceive to be (vi) exclusionary and (vii) agent-centered.[102] These refinements will be explained as they are introduced. They should be understood as deep structural features of obligata, and, hence, of morality and law. This section will nevertheless begin by using the more minimal definition as set forth in elements (i)-(iii).

The term "obligatum" may sound odd at first, and perhaps even overly technical, but it resonates with two terms that will help to clarify what exactly is being proposed. First, the term "obligatum" resonates with the adjective "obligato," which means indispensable, or not to be left out,[103] as in an obligato accompaniment, which is an integral part of a larger musical performance. For reasons that will become clear below, the idea that each part of an obligatum is an integral and indispensable part of the functioning whole is an idea that the term "obligatum" should vividly connote. Second, the term "obligatum" brings to mind the idea of an obligation. This is also an appropriate connotation because obligata are—on the view elaborated here—part of the normal psychological background in which obligations can be said to arise and exist. Obligata are in fact the very attitudes we employ when we sustain conventions, the phenomena that David Hume famously described as follows:

> When [a] common sense of interest is mutually express'd, and is known to [all], it produces a suitable resolution and behaviour. And this may properly enough be call'd a convention or agreement betwixt us, tho' without the interposition of a promise; since the actions of each of us have a reference to those of the other, and are perform'd upon the supposition, that something is to be perform'd on the other part.[104]

Hume thought that "conventions" of this kind animated our sense of justice and moral and political obligation.[105]

Moving on to substance, there are two likely sources of resistance to the idea that we might have internal functional states with the properties of obligata. First, there is the problem already canvassed, concerning why, absent a rational basis, we would ever have the suppositions that go into obligata, or suppose that they will form part

---

[102] A requirement is typically called "agent-centered" if in at least some circumstances it purports to give each agent a different aim or goal, namely that *he* or *she* fulfill a given requirement even if by failing to do so he or she could cause two or more others to fulfill the requirement in equally weighty circumstances. *See, e.g.*, DEREK PARFIT, REASONS AND PERSONS, at 55; *see also* Stephen Darwall, *Agent-Centered Restrictions from the Inside Out*, 50 PHIL. STUD. 291-319, *reprinted in* DEONTOLOGY 112, 112 (Stephen Darwall ed., 2003).

[103] THE AMERICAN HERITAGE DICTIONARY OF THE ENGLISH LANGUAGE 1211 (4th ed. 2000).

[104] DAVID HUME, A TREATISE OF HUMAN NATURE 490 (Oxford Univ. Press 2d ed. 1978) (1739/1740).

[105] *Id.* at 477-84 (sense of justice), 501-25 (sense of moral obligations) 534-49 (sense of political obligation).

of a mutually concordant system with others' motives and suppositions. This might seem to involve a coincidence bordering on the miraculous. Second, there is a basic plausibility issue with the claim that we share attitudes of this particular kind. Obligata motivate us to act in self-sacrificing manners, and include suppositions that others will act in a similar way. It might seem implausible, at best, and tragically naïve, at worst, to think that nature would endow us with unsecured suppositions of others' self-sacrificing behavior. This implausibility is heightened when we reflect on the fact that there are numerous well-known difficulties understanding how altruistic or self-sacrificing traits might evolve and persist in the natural world.[106]

Fortunately, we will not be forced to countenance these problems if we approach the issue from a contemporary perspective, as illuminated by a number of recent theoretical developments. Modern work on convention, beginning with that of Thomas C. Schelling,[107] John Nash[108] and David Lewis,[109] has, for example, helped show how certain kinds of coordinated attitudes like perceptions of salience can be critical for the resolution of strategic game-theoretic situations that bear important similarities to the social contract problems under discussion here. There has also been a great deal of recent work on the evolution of altruism and our sense of justice, including work by a number of philosophers such as Philip Kitcher,[110] Allan Gibbard[111] and Elliot Sober,[112] to name a few. The burgeoning field of evolutionary psychology has also offered a way of approaching the mind that allows for naturalistically sound functional categorizations, and for the identification of real psychological capacities, that might otherwise elude philosophical and empirical inquiry—a fact that is being increasingly acknowledged within the philosophy of mind. And finally, new work by evolutionary game-theorists, such as Brian Skyrms,[113] Phil Kitcher[114] and Robert Sugden,[115] has helped to show how a

---

[106] *See, e.g.*, Philip Kitcher, *The Evolution of Human Altruism*, 90 J. PHIL. 497, 497 (1993) ("The problem of altruism has loomed large in evolutionary biology ever since Charles Darwin. How do tendencies to kindly, even self-sacrificial, behavior evolve in an unkind, Darwinian world?"); Kitcher, *Psychological Altruism*, *supra* note 71, at 288; Allan Gibbard, *Human Evolution and the Sense of Justice*, 7 MIDWEST STUD. PHIL. 33-34 (P.A. French et al. eds., 1982); Neven Sesardic, *Recent Work on Human Altruism and Evolution*, 106 ETHICS 128, 128 (1995) ("[It is common to believe that] inveterately altruistic creatures have a pathetic tendency to die before reproducing their kind.").

[107] *See* THOMAS C. SCHELLING, THE STRATEGY OF CONFLICT (1960).

[108] *See* John F. Nash, Jr., *The Bargaining Problem*, 18 ECONOMETRICA 155-162 (1950).

[109] *See* DAVID LEWIS, CONVENTION (Blackwell Publishers 2002) (1969).

[110] *See, e.g.*, Kitcher, *The Evolution of Human Altruism*, *supra* note 106, at 497-516; Kitcher, *Psychological Altruism*, *supra* note 71, at 283-316.

[111] *See, e.g.*, ALLAN GIBBARD, WISE CHOICES, APT FEELINGS: A THEORY OF NORMATIVE JUDGMENT (1990); Gibbard, *supra* note 106, at 31 46.

[112] *See, e.g.*, Elliott Sober, *What Is Evolutionary Altruism?*, 14 CAN. J. PHIL. 75-99 (1988); Elliot Sober, *Did Evolution Make Us Psychological Egoists?*, *reprinted in* FROM A BIOLOGICAL POINT OF VIEW 8-27 (1994).

[113] *See e.g.*, BRIAN SKYRMS, EVOLUTION OF THE SOCIAL CONTRACT (1996); Brian Skyrms, *Darwin Meets* The Logic of Decision*: Correlation in Evolutionary Game Theory*, 61 PHIL. SCI. 503-28 (1994).

[114] *See, e.g.*, Philip Kitcher, *The Evolution of Human Altruism*, *supra* note 106, at 497-516; Philip Kitcher, *Psychological Altruism*, *supra* note 71, at 283-316.

look at our pasts as presenting us with repeated game-theoretic problems can help to explain the evolution and persistence of various features of our moral psychologies—even some that are apparently altruistic. These and related developments have helped pave the way for the claims that will be defended here, though the claims themselves will be new and are particular to this account.

Let us begin, then, with a look at evolutionary psychology—a field that would appear at first blush to be oblique to the project of illuminating moral phenomena but that will ultimately prove to yield several unexpected and fruitful insights.[116] Recent work in evolutionary psychology has helped establish the fecundity of viewing the mind as a bundle of domain-specific and content-laden mechanisms, which sometimes interact with one another and/or with cues from the social or material environment in highly complicated ways, but that are ultimately finely tuned to solve specific and longstanding adaptive problems that recurred in our environment of evolutionary adaptation.[117] Adaptations of this kind are generally species-typical and often universal to normal members of the species.[118]

The most important point to recognize for the present purposes is that an adaptationist approach to the mind can allow for the identification of mental phenomena by function, and without the need of imposing the assumption that all of our attitudes must be separated neatly by their direction of fit.[119] When this basic intuition is connected up with the facts about obligata that have already been canvassed, the claim that we employ obligata to solve social contract problems can be exposed as the unexpected beneficiary of plausibility considerations arising out of a naturalistic approach to the mind. There would, after all, be clear adaptive advantages to a capacity to solve many social contract problems, and, hence, natural selection could explain how

---

[115] *See, e.g.*, Sugden, *supra* note 88, at 73-99.

[116] For a definitive collection of essays defending the basic adaptationist framework employed in modern evolutionary psychology and presenting numerous suggestive applications, see THE ADAPTED MIND, *supra* note 38.

[117] *See, e.g.*, Tooby & Cosmides, *supra* note 38, at 49-123 (explaining paradigm and collecting evidence).

[118] Although Cosmides and Tooby often speak as if adaptive traits must be universal in the sense that nearly every member of a species must have the trait, *see id.* at 64, evolutionary dynamics can, in fact, be used to explain not only so-called "monomorphic" adaptions but also so-called "polymorphic" adaptations. This distinction arises because some traits have adaptive values that are frequency dependent, in the sense that their adapative value depends on how many other members of the group have the same trait. Hence, evolutionary dynamics can sometimes drive a species to so-called "polymorphic" outcomes in which a portfolio of distinguishable adaptations arise and remain stable in distinct ratios. *See, e.g.*, PAUL E. GRIFFITHS, WHAT EMOTIONS REALLY ARE: THE PROBLEM OF PSYCHOLOGICAL CATEGORIES 62-64 (1997). This possibility need not be canvassed here because the structural features of obligata that will be discussed do not appear polymorphic. The issue of whether there are nevertheless polymorphic aspects to our moral and legal psychologies is an interesting one, but one that goes beyond the scope of this Article.

[119] The term "direction of fit" is used in the sense that John Searle, among others, has made familiar. *See, e.g.*, JOHN R. SEARLE, INTENTIONALITY 7-9 (1983). Searle believes that most representational mental states can be divided into two main categories, based on their direction of fit. Beliefs are said to have a "mind-to-world" direction of fit because their function is to represent things in a way that matches facts about the world. *See id.* at 8. Desires are said to have a "world-to-mind" direction of fit because their function is to represent non-actual states of affairs and to bring about changes in the world to make them actual in at least some circumstances. *See id.* at 7-8.

obligata might proliferate through a population, as an adaptation that functions by allowing its bearers to reap the cooperative benefits of these resolutions. The last section suggested, moreover, that creatures set up to await adequate evidence of the relevant commitments would spend their lives waiting; whereas creatures endowed with obligata, and with their peculiar kind of suppositions, which run ahead of the evidence, could begin to reap the benefits of cooperation. This in itself provides a plausible naturalistic explanation for why we might expect obligata to exist, if we can in fact resolve social contract problems without security mechanisms. And—in fact—the evidence suggests that we do.[120]

At the same time, an evolutionary explanation for the emergence of obligata would entail that the suppositions that go into them have precisely the dual kind of responsiveness and recalcitrance to evidence that we often see in our normative practices—and that can otherwise seem so puzzling. An unconditional motivation to fulfill one's part of a social contract would incline its bearers to act in self-sacrificing ways for the benefit not only of others who are similarly inclined but also of those who have no such inclinations, thus endowing non-cooperators with selective benefits that would, all else being equal, outweigh those of the cooperators. There would thus be strong selective pressures against such unconditional motivations, and the proliferation of obligata through a population would seem to require some responsiveness to evidence of whether others are cooperating and some kind of channeling of the self-sacrificing behavior toward other fellow cooperators. The suppositions should, in other words, involve a kind of default trust, but a kind that is nevertheless defeasible and responsive to evidence of insufficient motivation.

But there should also be a corresponding ambiguity in the experience of a failed supposition. Such an experience might be taken as (i) challenging the idea that all (or a significant majority of) other relevant cooperators agree with our conception of the right or of what is required—which might thus invite normative discussion, with the aim

---

[120] It is common to observe that human relations give rise to numerous social contract problems. *See, e.g.*, Gauthier, *supra* note 82, at 98; SHAUN P. HARGREAVES HEAP & YANIS VAROUFAKIS, GAME THEORY 149-55 (1995). John Rawls calls these circumstances the "circumstances of justice" and takes them to be integral to much of our social and political lives. *See* RAWLS, *supra* note 65, at 127-28. Alan Fiske has been developing a comprehensive theory of the basic modes in which we conduct our social life and relations with one another, and all of which he considers to be modes of resolving problems of cooperation. *See, e.g.*, ALAN PAGE FISKE, STRUCTURES OF SOCIAL LIFE (1993).

Moreover, it would be difficult to explain how we might resolve social contract problems by creating sanctions that function as security mechanisms, unless we had some capacity to resolve them without such mechanisms. This is because systems of punishment and sanction are themselves typically public goods, which thus also have the underlying structure of an *n*-person prisoners' dilemma or social contract problem. Security mechanisms such as these cannot fundamentally explain how we resolve social contract problems, then, on pain of regress. For a good description of this problem, see Allan Gibbard, *Norms, Discussion, and Ritual: Evolutionary Puzzles*, 100 ETHICS 787, 797 (1990); *see also* Elizabeth Anderson, *Beyond* Homo Economicus: *New Developments in Theories of Social Norms*, 29 PHIL. & PUB. AFF. 170, 182-84 (2000). While a number of people have proposed resolutions to this problem, no proposal has won widespread acceptance. *Compare, e.g.*, Gibbard, *supra*, at 798 *with* Anderson, *supra*, at 181 (citing JON ELSTER, THE CEMENT OF SOCIETY 132-33 (1989)); *see also* Philip Pettit, Virtus Normativa: *Rational Choice Perspectives*, 100 ETHICS 725-55 (1990).

of settling on common ground;[121] as (ii) evidence that the person who was supposed to perform the relevant commitment is insufficiently motivated;[122] or, finally, as (iii) evidence that the supposed "other" is not a candidate for full participation in the social contract.[123] When placed in tandem with our other evidence and beliefs, however, which might help to disambiguate things a bit, these suppositions might nevertheless help us identify these important situations.

To mark the ways in which the suppositions that go into obligata both run ahead of the evidence and are in some ways recalcitrant to the evidence, it will be useful to label the suppositions "normative suppositions." These considerations provide the first promised refinement to the definition proposed thus far—as set forth in element (iv)—and clarify one set of ways in which these expectation-like states differ from ordinary belief-like states and involve certain attitudes of unsecured, default trust. Normative suppositions that function in these ways are part of the deep structure of morality and law.

It is, moreover, important to recognize, even at this early stage of the argument, that the various elements that go into obligata must be bound up together for them to serve their proposed function. For consider the alternatives. If we were to have the conditional motivations without the corresponding normative suppositions, then the resolution of the social contract problem would never get off the ground because we would never suppose the conditions to obtain that would call for us to fulfill our individual obligations. Similarly, if we were to have the relevant normative suppositions without the conditional motivations, none of us would be motivated to fulfill our individual obligations in a committed way. Our normative suppositions would thus be systematically invalidated, and the domain of others to whom we would deem ourselves obligated would narrow in scope to none. Finally, if we were each to have the conditional motivations and were to suppose all others to have them too, but were not to suppose that all others suppose all others to have them—a mouthful, indeed!—then we could not sustain the normative supposition that all others take themselves to be in the

---

[121] *See generally* Sections G and H, *infra* (elaborating how normative discussion works to produce coordination over normative content).

[122] *See generally* Sections D and E, *infra* (elaborating how we sometimes take deviations from norms as evidence of non-cooperative motive and how we tend to react to such conclusions).

[123] As Peter Railton has noted: "It is a commonplace of anthropology that tribal peoples often have only one word to name both their tribe and 'the people' or 'humanity.' Those beyond the tribe are not deemed full-fledged people, and the sorts of obligations one has toward people do not fully apply with regard to outsiders." Railton, *supra* note 78, at 27 (noting also that modern morality is viewed as extending to nothing short of the species). Gilbert Harman has similarly traced out a number of features of our ordinary moral judgments that suggest we sometimes view people in different cultures or who are so far removed from our basic moral outlook as improper objects for sincere moral claims. *See, e.g.*, GILBERT HARMAN, THE NATURE OF MORALITY (1977), *reprinted in* CONTRACTARIANISM/ CONTRACTUALISM, *supra* note 67, at 140-144. The same is often true in the law, where we typically view foreign persons to be bound by their domestic laws, not ours—at least in most circumstances. There is, moreover, a great deal of psychological literature confirming that we tend to cognize many social relations in terms of "in-group"/"out-group," and that the presence or absence of shared group identification has palpable effects on our levels of cooperative motive. *See, e.g*, Richard H. McAdams, *Cooperation and Conflict*, 108 HARV. L. REV. 1003, 1014-16 (1995); Naomi Ellemers & Wendy Van Rikswijk, *Identity Needs Versus Social Opportunities*, 60 SOC. PSYCHOL. Q. 52, 52 (1997).

circumstances needed for their obligations to make claims on their individual conduct—and the conventional enterprise would tend to unravel. Hence, just like an obligata accompaniment of a larger musical performance, the various elements that constitute obligata, and that remain fixed to a common content, must come together in a unitary bundle if they are to serve their proposed function.

The initial step in recognizing the existence of obligata is thus to acknowledge, first, that we sometimes solve social contract problems without security mechanisms,[124] and, second, that given our best naturalistic understanding of ourselves, plausibility considerations speak in favor of our employment of these strange bundles of coordinated motive and normative supposition—or at least something very much like them—to achieve this otherwise ordinary social task.

D. ARGUMENT FROM STABILITY CONDITIONS AND REACTIVE ATTITUDES

The arguments presented thus far should not produce conviction independently. They importantly do more than merely suggest that we have capacities to resolve social contract problems based on the fact that these capacities would be beneficial. They note, instead, that we seem to resolve a variety of social contract problems quite naturally and then suggest on this basis, and on the basis of what we would need to do this, that we likely have psychological capacities with a particular structure. Still, these initial arguments are only preliminary, and their deeper importance lies in how they set the stage for a much more decisive set of arguments for the existence and function of obligata as proposed herein.

The method that will be used to illuminate these more decisive grounds is simply to hypothesize the existence of obligata as adaptive attitudes that function by allowing us to resolve social contract problems, and then to ask when such attitudes would be evolutionary stable. The answer to this question will generate empirical predictions about some of the other motivational and behavioral phenomena that we should expect to accompany any genuine attitudes of this kind. An examination of our moral and legal practices, in the context of the larger ethnographic and psychological record, as well as a number of independently derived philosophical accounts of obligation, can then be used to test the plausibility of the current proposal. This Section turns to a core part of this task.

As it turns out, this methodology will be quite useful because obligata, as thus far defined, appear to have a particular property, namely that of being evolutionarily altruistic. Following Elliot Sober, a trait will be called "evolutionarily altruistic" if it inclines an organism to act in ways that enhance the reproductive fitness of one or more other organisms at some cost to the reproductive fitness of the individual bearing the trait.[125] Obligata might seem to do this because they function to resolve social contract

---

[124] *See supra* note 120 (collecting citations observing that human life is replete with social problems) (noting that we must have a capacity to resolve such problems without security mechanisms).

[125] Sober says that "[i]n evolutionary biology, . . . the concept [of altruism] is applied to behaviors that enhance the fitness of others at the expense to self." Elliot Sober, *Did Evolution Make Us Psychological Egoists?*, *supra* note 112, at 8.

problems, which have the underlying structure of an *n*-person prisoners' dilemma,[126] and because they therefore motivate behavior that is strictly dominated (in the language of rational choice theory).[127] Where the resolution of a social contract problem has adaptive value, any ordinal payoff relations expressed by the rational choice theorist can be rendered in terms of ordinal fitness consequences, and the motives to cooperate will turn out to have evolutionarily altruistic properties as well. Now, there is a well known problem with the stability of evolutionarily altruistic traits. All other things being equal, the bearers of such traits will tend to do poorer in evolutionary time than their more selfish counterparts. If applicable to obligata, this would mean that any bearers of obligata with motives of a given strength would tend to do more poorly than any counterparts with lesser, but non-neglible, motives. Over generations, the members who were less motivated by the social contract would tend to do better, and obligata with any notable strength would tend to slowly erase themselves from the natural world. This is a well-known phenomenon that is commonly referred to as "subversion from within."[128] And the only constraint that has been placed on this subversion thus far is that a complete absence of cooperative motivation will keep one from being a candidate for membership in the social contract, and for the benefits that naturally flow therefrom. This constraint still allows for the subversion of obligata with any notable strength, and, hence, for the subversion of many if not most resolutions employing this psychological state.

There is, however, now a large body of literature on the conditions under which traits with seemingly evolutionarily altruistic properties can evolve and remain stable against subversion. Through the modern study of "replicator dynamics," we can model populations of self-replicating organisms that interact with one another and with the world in certain well-defined recurrent situations, such as those having the structure of a social contract problem, in accordance with a set of competing traits that define how they will act in the given interactions.[129] The expected fitness "payoffs" from the modeled interactions can then be calculated, and relative payoffs can be used to determine the percentage of "offspring" that each organism will leave to interact in

---

[126] The standard prisoners' dilemma is a two-person a game. For a clear introduction to the standard example, see Shaun Hargraves Heap, *Game Theory*, *in* HEAP ET AL., *supra* note 78, at 98-100. Prisoners' dilemmas can, however, arise among larger groups of *n* persons, and an "*n*-person prisoners' dilemma" refers to any *n*-person game in which there is a cooperative equilibrium that is strictly dominated by an alternative defecting equilibrium, and where rational utility maximizers would therefore lose the potential cooperative benefits of resolving the dilemma if they were to act in accordance with this norm of rationality.

[127] Rational choice theorists call a strategy "strictly dominated" if it is "never as good as another feasible strategy, whatever the other player does." Bruce Lyons, *Game Theory*, *in* HEAP ET AL., *supra* note 78, at 98. Acting in accordance with a cooperative equilibrium in a prisoners' dilemmas is strictly dominated in this sense. *See id.*

[128] *See* Samir Okasha, *Biological Altruism*, *in* STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2003), http://plato.stanford.edu/entries/altruism-biological (noting use of this term and describing phenomenon); *see also* RICHARD DAWKINS, THE SELFISH GENE 7-8 (new ed. 1989) (presenting general argument that natural selection should be expected to weed out altruistic traits).

[129] For a useful introduction to replicator dynamics and evolutionary game theory, see Peter Hammerstein, *What is Evolutionary Game Theory?*, *in* GAME THEORY AND ANIMAL BEHAVIOR 3-15 (Lee Alan Dugatkin & Hudson Kern Reeve eds., 1998).

accordance with its given trait in the next generation.[130] The question whether a given trait, with certain seemingly evolutionarily altruistic features, could evolve and persist as an adaptation given the entry of certain near variants can then be reframed as the question whether a pure population with that trait is a "Nash equilibrium" in the replicator dynamics. To say that something is a "Nash equilibrium" is to say that interacting in accordance with the particular equilibrium trait in a population dominated by that same trait yields higher fitness benefits in that population than any of the other variants that are modeled in the replicator dynamics.[131] There can be multiple Nash equilibria, but any strategy that meets this criterion is said to be "evolutionarily stable."[132]

---

[130] It is common to begin by saying that, in the initial state of a system, there are *n* different strategies s1, . . . , s_n, which reflect different ways the bearers of the traits will interact in the relevant game theoretic situations, and to denote the frequency of strategy S_i as f_i. *See, e.g.*, *id.* at 5. Often, the relative payoffs associated with playing the various strategies will be independent of the frequencies of the strategies. This is not, however, always the case, and it will not typically be the case for the problems discussed here. To capture the possible frequency-dependent aspects of these payoffs, it is common to define a vector that represents the relative frequencies of each strategy, such as $f = (f_1, \ldots, f_n)$, and then to define the expected fitness associated with playing strategy S_i as $u_i(f)$. The mean expected fitness value of the population can, finally, be represented as $\bar{u}(f)$. Then the vector representation of the relative frequencies of each strategy in a subsequent generation can be represented as $f' = (f'_1, \ldots, f'_n)$, and the value can be calculated with the following equation (which is often called the "discrete replicator equation with frequency-dependent fitness"):

$$f'_i = f_i \frac{u_i(f)}{\bar{u}(f)} \qquad \text{for } i = 1, \ldots, n$$

Id. Quite often, iteration of this equation will yield one or more "Nash equilibria," or vector states that stabilize themselves and resist perturbation by increases or decreases of competing strategies once the vector state has been reached. By applying this equation to a number of game theoretic that one might plausibly think recurred in our environment of evolutionary adaptation, one can thus get a sense of which strategies might have won out against others, beginning either with random populations or more limited sets that reflect facts that we learn about our natural histories. The percentage of random populations that statistically end up at one or another Nash equilibrium is commonly called the "basin of attraction" for that equilibrium, and Nash equilibria with larger basins of attraction are, all other things equal, more likely evolve in nature. *See generally* SKYRMS, *supra* note 113, at 14-16, 19-21 (describing basins of attraction and how they operate in application to specific game theoretic problems).

The reader will be happy to know that the discussions in the main text will not require the level of mathematical detail just canvassed.

[131] *See, e.g.*, Lyons, *supra* note 127, at 101 ("A Nash equilibrium is a set of strategies, one for each player, such that given the strategies being played by others, no player can improve on her pay-off by adopting an alternative strategy. This concept is so fundamental that it is often called simply the equilibrium point.").

[132] The concept of an "evolutionary stable strategy" was first introduced by John Maynard Smith and G.R. Price in *The Logic of Animal Conflicts*, 24 ANIMAL BEHAVIOUR 159-175 (1973). There are some variations in how the term is used. Allan Gibbard says that an evolutionarily stable strategy is any "strategy such that given the organism's environment, which consists in part of the behavioral dispositions of other organisms, its strategy is at least as fitness enhancing as any other strategy easily accessible by mutation." Gibbard, *supra* note 106, at 31, 35. Maynard Smith uses the term in a somewhat stronger sense and says that an evolutionary stable strategy "is a strategy such that, if all members of a population adopt it, then no mutant strategy could invade the population under the influence of natural selection." JOHN MAYNARD SMITH, EVOLUTION AND THE THEORY OF GAMES 10 (1982). Gibbard's definition would add a useful dose of realism and empirical plausibility to many applications of evolutionary game theoretic reasoning, but nothing important hinges on these distinctions for present purposes.

Notice that any trait meeting this criterion will not genuinely be evolutionarily altruistic, in the final analysis, because—by some complex method or other—the trait must be more advantageous to its bearer than any relevant alternative if it is to be evolutionarily stable. Still, standing alone, the primary motives to act in accordance with standards that resolve social contract problems are evolutionarily altruistic for reasons already discussed; we must therefore look for other features of these attitudes to identify what would stabilize them and make them lose this property in that more complex form. When bound up with these other psychological phenomena, our motives to act in accordance with morality and law would, however, still have a related property that can be puzzling from an evolutionary perspective. They would be altruistic in an *adaptationist* sense, which Cosmides and Tooby have recently clarified as follows:

> An adaptationist definition of altruism would focus on whether there was a highly nonrandom phenotypic complexity that is organized in such a way that it reliably causes an organism to deliver benefits to others, rather than on whether the delivery was costly. The existence of such a design problem is the adaptationist problem of altruism—an evolutionary "problem" requiring explanation whether that delivery is costly, cost-free or even secondarily beneficial to the deliverer.[133]

The project here, then, is to identify evolutionary stability conditions for seemingly evolutionarily altruistic parts of obligata, which conditions will reveal that obligata as a whole are ultimately adaptive and are not evolutionarily altruistic. But to reach this conclusion we must work through the puzzle of how obligata, as so far defined, might arise and persist in nature, given that they seem to be evolutionarily altruistic.

Through the study of replicator dynamics, it has been shown that there are, in fact, a number of discrete processes that can produce such traits. These are the processes of kin selection,[134] identification and discrimination,[135] certain highly-specific forms of geographical clustering forced by external circumstances,[136] reciprocal altruism[137] and, arguably, certain forms of so-called "non-naïve" group selection that ultimately depend upon mechanisms like highly-specific forms of geographical clustering

---

[133] John Tooby & Leda Cosmides, *Friendship and the Banker's Paradox: Other Pathways to the Evolution of Adaptations for Altruism*, 88 PROC. BRIT. ACAD. 119, 121 (1996).

[134] W.D. Hamilton is typically credited with clarifying that survival of the fittest operates on *inclusive* fitness, which is "measured by its effect on survival and reproduction *both of the organism bearing it, and of the genes, identical by descent, borne by the organism's relatives.*" DOUGLAS FUTUYAMA, EVOLUTIONARY BIOLOGY G-1 (3d ed. 1998) (emphasis added). This fact can be used to explain the evolution of traits that are seemingly evolutionarily altruistic toward kin, and the mechanisms that produce such traits have been called "kin selection." *See, e.g.*, DAWKINS, *supra* note 128, at 89-108.

[135] *See, e.g.*, Kitcher, *The Evolution of Human Altruism*, *supra* note 110, at 497-516.

[136] *See, e.g.*, Sober, *Did Evolution Make Us Psychological Egoists?*, *supra* note 112, at 8-27.

[137] *See, e.g.*, ROBERT AXELROD, THE EVOLUTION OF COOPERATION (1984) (developing the concept of "reciprocal altruism" and using computer modeling to show how certain forms of reciprocal altruism can evolve and remain stable in nature).

or identification and discrimination.[138] In canvassing these and related possibilities, Brian Skyrms has recently observed that *positive correlation* between evolutionary altruists is the common feature that allows evolutionary forces to select for phenomena of this kind.[139] Although the mechanisms that ensure this correlation can be quite varied, the correlation is what is mathematically significant in the replicator dynamics, and is what gives all the mechanisms that we currently understand their ability to generate and sustain cooperative equilibria.[140]

The reason that positive correlation captures something important in the replicator dynamics is, however, that positive correlation helps ensure that any benefits of a cooperative enterprise flow primarily to cooperators. A positive correlation is, in fact, sufficient to allow for the evolution of cooperation only if the increased benefits that cooperators obtain from their cooperative efforts due to the positive correlation are larger than both the costs involved with cooperating and the benefits, if any, that non-cooperators also obtain from the cooperative enterprise. But this suggests that what is fundamental is not positive correlation itself but rather this relational property concerning the distributions of evolutionary costs and cooperative benefits among cooperators and non-cooperators. Where this distribution is not guaranteed by kin selective forces or by mechanisms external to the group, this means that a basic stability condition of obligata is that cooperators must share internal psychological mechanisms or capacities to identify and exclude non-cooperators from the benefits of the cooperative enterprise, either by preventing sufficient benefits from flowing to non-cooperators or by engaging in precommitted and costly acts of punishment that will make non-cooperation sufficiently costly. Obligata could remain evolutionarily stable, then, if—as set forth in element (v) of the refined definition—the breach of normative suppositions were to trigger emotions or other powerful impulses or attitudes that would function to identify and exclude non-cooperators from the benefits of the social contract.

Once this evolutionary stability condition has been acknowledged, a

---

[138] Group selectionist theories have generally fallen into disfavor, and are likely inconsistent with natural selection. Most commentators agree that at minimum "the conditions necessary for [group selection] to occur are quite stringent," and the phenomenon is "quite rare"—if it exists at all. David L. Hull, *Introduction to Part III: Units of Selection*, *in* THE PHILOSOPHY OF BIOLOGY 149, 149 (David L. Hull & Michael Ruse eds., 1998). The leading modern proponents of a so-called "non-naïve" group selectionist account of human altruism are Elliot Sober and David Sloan Wilson. *See, e.g.*, ELLIOTT SOBER & SLOAN WILSON, UNTO OTHERS: THE EVOLUTION AND PSYCHOLOGY OF UNSELFISH BEHAVIOR (1998). Sober and Wilson have identified a discrete set of circumstances in which so-called "non-naïve" group selection purportedly operates. The circumstances in question are, however, ones that meet the more general criterion for individual or gene-level selection to produce and stabilize traits with evolutionarily altruistic properties that will be developed in the main text of this Article below. *See infra* note __ & accompanying text. For the purposes of this Article, there is no reason to distinguish these circumstances from ordinary instances of natural selection operating, and use of the term "group selection" to refer to these circumstances here may actually lead to confusion.

[139] *See* SKYRMS, *supra* note 113, at 61. For a more in depth treatment of these issues, which shows how positive correlation arises in these other settings, see Skyrms, *supra* note 113, at 503-28.

[140] *See, e.g.*, *id.* at 525 (noting that correlated interactions "may be a consequence of a tendency to interact with relatives (Hamilton's kin selection), of identification and discrimination, of spatial location, or of strategies established in repeated game situations (the reciprocal altruism of Trivers 1971 and Axelrod and Hamilton 1981)").

wealth of considerations can be seen to speak in favor of the existence of obligata with their proposed function. Beginning with the evidence from moral philosophy, Stephen Darwall has, for example, recently emphasized that we cannot even begin to understand the distinction between the normativity of moral obligations and those of other purportedly categorical requirements, such as the requirements of logic or etiquette, without conceding an intrinsic relation between obligations and others' standing to demand compliance, sometimes by invoking reactive attitudes or other forms of permitted punishment or coercion for non-compliance.[141] This kind of concession would seem to entail that our attitudes and practices relating to moral obligation are inherently bound up with the kinds of phenomena needed to stabilize obligata. In more or less explicit form, this same kind of relation has, moreover, been recognized by a number of previous philosophers reflecting merely on the structure of our moral practices and/or on the meaning of our moral terms. Richard Brandt has, for example, listed among the criteria by which a sociologist might recognize the 'moral code' of a society, not only that (i) individuals have intrinsic motivations to respect the relevant moral imperatives but also that (ii) individuals "think[] it proper that some degree of coercion be brought on a person (perhaps only by the pressure of his own conscience) to induce the relevant form of behaviour in him" and that (iii) actions contrary to the code are met with "guilt-feelings and disapproval."[142] John Stuart Mill has similarly analyzed moral obligation in terms of moral wrongdoing, stating, famously, that: "We do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience."[143] And attitudes like these are, in fact, nothing other than the reactive attitudes that Peter Strawson has famously shown to be central to our ordinary concept of moral responsibility[144] and—though the point is less frequently noted—of moral obligation as well.[145] All of these thinkers reached their respective conclusions without the aid of the kind of evolutionary argumentation presented here, and the coincidence of their results with the claims defended here on independent naturalistic and theoretical grounds suggests that both sets of views do great justice to the truth.[146]

---

[141] *See, e.g.*, Darwall, *supra* note 21, at 129, 136-38, 144-53.

[142] BRANDT, *supra* note 5, at 163-70.

[143] JOHN STUART MILL, UTILITARIANISM 193 (Bantam Books 1993) (1863).

[144] For the classic exposition of this view, see P.F. Strawson, *Freedom and Resentment*, *in* STUDIES IN THE PHILOSOPHY OF THOUGHT AND ACTION (P.F. Strawson ed., 1968).

[145] In explanation of our tendency to inhibit the reactive attitudes towards persons who are deemed insane, Strawson says, for example, that "to the extent to which the agent is seen in this light, he is not seen as one on whom demands and expectations lie in that particular way in which we think of them as lying when we speak of *moral obligation*; he is not, to that extent, seen as a morally responsible agent, as a term of moral relationships, as a member of the moral community." *Id.* at 88 (emphasis added).

[146] Drawing on a number of closely related evolutionary considerations, Allan Gibbard has also accounted for what he calls "narrow" moral judgments as expressing attitudes of norm acceptance that govern attitudes of guilt and impartial anger. *See* GIBBARD, *supra* note 111, at 23-82. Gibbard's account posits a psychology that is, in its basic contours, fully in line with the main thrust of the psychological claims defended here. Though our psychological accounts differ in a number of important details, our views converge significantly, at least at this level of abstraction. Many of Gibbard's arguments and those

In any event, further support for the claims defended here can be found if we step back from accounts of moral obligation and look to related accounts of legal obligation. In *The Concept of Law*, H.L.A. Hart has famously argued that when we use the normative language that pervades our legal practices, we are typically giving expression to rules that we accept "internally." By this, he means that we not only take the rules as internal guides to action but also view their breach as warranting criticism of some sort.[147] When we take up an internal attitude to moral and legal obligations, he later says, moreover, that the commands they present us with "may be taken not only as [i] a peremptory [exclusionary] guide to action by those who are themselves commanded to act, but may be taken by them and others as [ii] a standard of evaluation of the conduct of others as correct or incorrect right or wrong . . . and as [iii] rendering unobjectionable and permissible what would normally be resented, that is demands for conformity, or various forms of coercive pressure on others to conform, whether or not those others themselves recognize the commands as peremptory reasons for their own actions."[148] These last attitudes, along with their well known connection to practices of enforcement and punishment, are just what would be needed to provide for the evolutionary stability conditions of obligata, conceived as adaptations that function by allowing us to resolve social contract problems.[149]

We can, finally, step back even further and find additional support for the claims made here in the larger ethnographic and psychological record. Christopher Boehm—a leading cultural anthropologist—has, for example, collected data on the social structures common to many hunter-gatherer tribes and bands, whose social lives most plausibly resemble ordinary human life during most of our environment of evolutionary adaptation.[150] His work suggests that such bands tend towards what he calls an "egalitarian ethos," with violations of egalitarian norms generating reactions of ridicule, ostracism, physical sanctioning, exile, and sometimes even group killings of norm violators.[151] Indeed, moralistic aggression and anger at norm violations is a seemingly

---

presented here should thus be viewed as reciprocally reinforcing with regard to the psychological claims defended.

[147] *See, e.g.*, HART, *supra* note 5, at 82-90, 242.

[148] H.L.A. Hart, *Commands and Authoritative Legal Reasons*, *reprinted in* AUTHORITY 103 (Joseph Raz ed., 1990).

[149] In distinguishing between questions about what we ought to do, all things considered, and what we have an "obligation" or "duty" to do, Ronald Dworkin similarly observes that "[j]udgments of duty are commonly much stronger than judgments simply about what one ought to do. We can demand compliance with an obligation or a duty, and sometimes propose a sanction for non-compliance, but neither demands nor sanctions are appropriate when it is merely a question of what one ought, on the whole, to do." DWORKIN, *supra* note 2, at 48. He continues that "[t]he question of when claims of obligation or duty are appropriate, *as distinct from such general claims about conduct*, is therefore an important question of moral philosophy, though it is a relatively neglected one." *Id.* (emphasis added).

[150] *See* CHRISTOPHER BOEHM, HIERARCHY IN THE FOREST 198 (1999) (noting that "later *Homo erectus* and Neanderthal apparently lived in smallish bands like those of extant mobile hunter-gatherers").

[151] *See, e.g.*, *id.* at 30 63, 214-15

cross-cultural feature of human society,[152] one which anthropologists have suggested persists even through otherwise vast cultural differences and differences in the social and political structures of societies.[153] Moving from anthropology to psychology, there is, moreover, now a growing body of evidence that we engage in costly forms of punishment for non-cooperation in many public goods and prisoners' dilemma situations.[154] Whether we approve of these facts about ourselves or not, they provide powerful support for the claim that we share obligata and that they serve the function proposed here. Notice, moreover, the form of the support: the account has yielded testable, empirical predictions, which cohere with and find support in a wealth of the available data.

Let us return, then, to element (v) of the proposed definition. This element posits mechanisms that function to identify and exclude non-cooperators from the benefits of the cooperative enterprise, which are triggered, consequent upon any breach of the normative suppositions. Viewed from an external standpoint, this is merely a causal claim. But viewed from the perspective of the bearers of obligata, who in bearing them take there to be reason to fulfill their respective obligations and to suppose others to do the same, this refinement corresponds to the psychological fact that these same bearers take there to be reason to react to deviations in certain ways and/or to permit what would

---

[152] *See, e.g.*, *id.* at 214-15, 245-47; Haidt, *supra* note 12, at 5-8; Fessler & Haley, *supra* note 12, at 12-14; Nucci & Nucci, *Children's Responses to Moral and Social Conventional Transgressions in Free-Play Settings*, *supra* note 52, at 1340.

[153] Another primitive and seemingly recurrent systems of norms that anthropologists know of are systems of honor and revenge. These are systems in which perceived breaches of norms are met with precommitted and costly impulses to avenge the perceived wrong, which impulses are often seconded by kin-selected impulses to cooperate, thus yielding the well-known phenomenon of the blood feud in which the perceived wronging of one person can generate a committed and concerted effort on the part of the entire person's family to avenge the initial perceived wrong. Such precommitted impulses are ubiquitous enough in pre-state societies that Jon Elster can claim that the motivation for revenge is "universal" in the human species, Jon Elster, *Norms of Revenge*, 100 ETHICS 862, 862 (1990), and Robert Frank can say that "[w]here the force of law is weak, cycles of attack and revenge are familiar. They pervade life in the Middle East today and have been recorded throughout human history." ROBERT H. FRANK, PASSIONS WITHIN REASON: THE STRATEGIC ROLE OF THE EMOTIONS 2 (1988). Such motivations persist even after the rise of modern states, however. The anthropologists Martin Daly and Margo Wilson have documented that the leading source of homicide in human societies cross-culturally is what they call "altercations of relatively trivial origin." *See* MARTIN DALY & MARGO WILSON, HOMICIDE 123-36, 221-52 (1988). These are situations in which a perceived insult or dishonor begins a cycle of precommitted responses that can escalate and ultimately end in bloodshed.

It is unclear whether honor codes engage obligata as these attitudes have thus far been described. The element of committed punitive reactions to perceived wrongs would seem to fit part of the model, but the attitudes that sustain honor codes also function very poorly in many familiar circumstances, including most modern ones. Honor codes may engage attitudes like obligata that functioned well for certain problems that recurred in our environment of evolutionary adaptation but that are systematically dysfunctional in many modern circumstances. One thing seems clear: absent attitudes of deference to external authority to adjudicate these kinds of potentially escalating conflicts, honor codes can lead to much that is wasteful and counterproductive.

[154] *See, e.g.*, Ernst Fehr & Simon Gächter, *Altruistic Punishment in Humans*, 415 NATURE 137-40 (2002); Ernst Fehr & Simon Gächter, *Cooperation and Punishment in Public Goods Experiments*, 90 AM. ECON. REV. 980-94 (2000). On the costs that are ordinarily involved in committed dispositions to punish, see, e.g., SKYRMS, *supra* note 113, at 22-28; FRANK, *supra* note 153, at 1-4.

otherwise be resented, namely, certain forms of punishment or coercion for non-compliance. These are features that should be familiar from many of our normative practices, and are—on the present view—parts of the deep structure of morality and law.

In addition, we can now specify the relationship between the perceived reasons to act that arise from obligata and those that arise from the broad class of desires assumed at the beginning this argument.[155] The function of obligata is to resolve social contract problems that arise in light of these ordinary desires. To fulfill this role, obligata must therefore provide us with sufficient motivation to counteract at least some of those desires. Corresponding to this motivational fact are several phenomenological ones concerning how obligations should appear to us in first person deliberation. First, we should expect that we would perceive obligations to have the standing to override or exclude at least some reasons arising from desire or personal interest. This is, in fact, precisely the standing that Raz has referred to in arguing that moral and legal imperatives provide us with "exclusionary" reasons to act.[156] Second, we should expect that we will perceive obligations as providing us with reasons that are in some sense irreducible to our antecedent desires or interests and arise from rules that apply to us regardless of antecedent desire or interest. These are, in fact, parts of the ordinary intuitions we have when we think that moral and legal obligations to have a peculiar binding nature or categorical force.[157]

Of course, it is probably more phenomenologically accurate to suggest that obligata function along with our other ordinary beliefs and desires in a way that renders action in accordance with morality and law less conscious and more habitual in many circumstances. For example, most of us probably do not even think about things like murdering people in order to obtain things that we desire. Obligata would nevertheless serve their function perfectly well if, when placed in conjunction with other psychological mechanisms of habituation, they were to provide us with what we took to be sufficient reasons to act. With these caveats in mind, we should nevertheless include the (vi) exclusionary force of the reasons that morality and law provide us with as part of the deep structure of morality and law.

Before continuing, it is, finally, worth pausing for a moment to clarify the sense in which obligata are and are not altruistic. The arguments thus far have suggested that while obligata are not strictly speaking evolutionarily altruistic, they do incline us to act in ways that regularly confer evolutionary benefits on others and to do so at what would—absent the stabilizing reactive attitudes—regularly induce evolutionary costs to ourselves. This does not yet speak to the question of whether obligata incline us to act in ways that are psychologically altruistic. The question whether an action is psychologically altruistic is one concerning the type of motive from which one acts, and whether—for example—one acts out of concern for others or to procure one's own interests, regardless of whatever evolutionary costs or benefits the act might otherwise

---

[155] *See* Section B, *supra*.

[156] *See, e.g.*, RAZ, *supra* note 1, at 16-25.

[157] *See, e.g.*, Brink, *supra* note 20, at 255-67, 280-87.

have.[158] In fact, however, many of the motives that go into obligata in us would appear to be neither psychologically altruistic nor psychologically selfish. They are better characterized as deontological attitudes, which allow us to take the perceived fact that something is right or required as a sufficient reason to act.[159] Acting on such reasons may require us to do things that we can understand, from another perspective, as involving self-sacrifice or as conducing to the benefit of others, but these considerations need not be our own reasons for action if we are instead acting out of a sense of duty. These points are important to bear in mind because much of the existing literature on the evolution of the moral sentiments is focused on the issue of altruism, and does not always appreciate the degree to which psychologically deontological, rather than psychologically altruistic, motives sustain our moral and legal practices.

E. ARGUMENT FROM THE STANDARD EXCUSES

At this point, a third set of considerations that speak in favor of the view that we use obligata to respond to obligations can be brought to light. The last section argued that phenomena like the reactive attitudes would allow natural selection to stabilize obligata with a given equilibrium strength of motivation. Hence, if we were to have well-functioning obligata, which give life to a system of obligations, we might expect to find some evidence in our normative practices of sensitivities to a particular distinction. This is the distinction between failures to act in accordance with directives that genuinely reveal insufficient motivation—as assessed by this equilibrium standard—and failures that do not. Only failures that genuinely reveal insufficient motivation should be deemed evidence of non-cooperation.[160] And while a sensitivity to this distinction is not a general evolutionary stability condition for obligata, there would be clear adaptive value to the sensitivity. It would allow the bearers of obligata to refrain from wrongly excluding genuine cooperators whose conduct nevertheless seems, on its face, to be in breach. Hence, if we see some sensitivity to this distinction in our normative practices——perhaps even a growing sensitivity in the course of human history—these facts will provide further grounds for the claims defended thus far.

Reflection on a number of general facts about us and the world we live in will, moreover, allow us to identify a number of regular and expectable situations that

---

[158] This definition of psychological altruism is thus the polar opposite of what so-called "psychological egoists" claim exhaust human action and motivation. *See generally, e.g.*, Joel Feinberg, *Psychological Egoism*, *in* REASON AND RESPONSIBILITY 493, 493 (Joel Feinberg & Russ Shafer-Landau eds., 10th ed. 1999) (defining "psychological egoism" as the view that we act only to further our own selfish interests or desires, and, hence, in terms of the motives from which we purportedly act) (noting that psychological egoism is "widely held by ordinary people, and [was] at one time almost universally accepted by political economists, philosophers, and psychologists").

[159] Whether there are any objective facts about right and wrong, and the like, or whether these terms are merely expressions of motivational states like obligata, is something that will be touched upon briefly below. *See* Section G, *infra.*

[160] In saying this, I do not mean to suggest that we primarily assess one another on a character-by-character basis rather than attributing responsibility to one another for particular actions. We clearly hold people responsible for many actions on an act-by-act basis. Still, when we do so, we are often highly sensitive to the quality of motive that went with the breach. *See, e.g.*, Strawson, *supra* note 144, at 74-91.

would cause any of us to fail to fulfill a directive without revealing a relevant motivational problem. These would include failures deriving from things like mistaken beliefs, the unforeseen consequences of our actions, the physical impossibility of fulfilling a commitment in a particular set of circumstances, momentary losses of control that would affect other ordinary persons just the same, and/or internal or external forces that simply outweigh the required equilibrium strength of motivations to comply. In fact, we are all familiar with practices of excuse-making that respond to just these situations. In our periodic recognition of the familiar excuses of mistake, accident, impossibility, incapacity, loss of control (sometimes dubbed irrationality or insanity), force, duress and necessity, we see sensitivities to the exact distinction in question.

This sensitivity is, moreover, pervasive in precisely the sense needed to support the claims in this Article. Richard Brandt has identified just these excusing conditions as playing a critical role not only in commonsense morality but in the criminal law;[161] and H.L.A. Hart has "draw[n] attention to the analogy between conditions that are treated by criminal law as *excusing* conditions and certain similar conditions that are treated in []other branch[es] of the law as *invalidating* certain civil transactions such as wills, gifts, contracts, and marriages."[162] Writing at a time that many believe to reflect ethical values that are in many ways deeply inconsistent with our own,[163] Aristotle nevertheless identified the very same set of excuses and argued that they are an intrinsic part of human ethical practices, which can be used to give content to the conception of responsibility or voluntariness that we use in these practices[164]—a tack that J.L. Austin took up much more recently, adverting, once again, to the very same excusing conditions.[165] Finally, while there is some debate as to whether ancient or so-called "primitive" legal and moral systems have allowed for these excuses—or have allowed for them to the same degree—some range in sensitivity is consistent with the claims developed here because our capacity to recognize the standard excuses contributes to the stability of our normative practices, on the present view, but is not a strict evolutionary stability condition. In any event, in his famous studies of primitive law, Oliver Wendell Holmes has suggested that even in the systems of vengeance, or laws of the blood feud, that characterized early Roman law and the laws of the Germanic tribes before the rise of the common law, only wrongful actions that were perceived as intentional were avenged;[166] and Frans de Waal, a prominent primatologist, has observed that a focus on distinguishing between deliberate and accidental actions is prominent even in the greater

---

[161] *See* Richard B. Brandt, *A Utilitarian Theory of Excuses*, *supra* note 6, at 215-16, 233; Brandt, *A Motivational Theory of Excuses in the Criminal Law*, *supra* note 6, at 235-38.

[162] Hart, *supra* note 6, at 29

[163] *See, e.g.*, ALASDAIR MACINTYRE, AFTER VIRTUE (1981); WILLIAMS, *supra* note 24.

[164] *See* ARISTOTLE, NICOMACHEAN ETHICS bk.III.1-3, 5, 10-12, at § 1110a-1113a, 1113b-1115a, 1117b-1119b (Martin Ostwald trans., Prentice Hall 1999).

[165] *See* J.L. Austin, *A Plea for Excuses*, 57 PROC. ARISTOTELIAN SOC'Y 1 (1956-57) *reprinted in* PHILOSOPHICAL PAPERS 175, 175-204 (2d ed., 1970). Strawson similarly accounts for moral freedom and responsibility in terms of reactive attitudes that are responsive to these same excusing conditions. *See* Strawson, *supra* note 144, at 74-91.

[166] *See* Oliver Wendell Holmes, Jr., *Early Forms of Liability*, *in* THE COMMON LAW I (1881), *reprinted in* THE ESSENTIAL HOLMES 237-38 (Richard A. Posner ed., 1992).

primate line.[167] Perhaps, then, there is some truth to Holmes's famous observation that "even a dog distinguishes between being stumbled over and being kicked."[168] If there is a distinction here between modern societies and so-called "primitive" societies, it would thus appear to be a matter of degree, not kind.

There is, however, an important objection that one might raise at this point. As described, the present account might seem to suggest that morality and law would limit liability to voluntary transgressions, but there are obvious counterexamples to this claim. Tort law is, for example, a full and thriving branch of Anglo-American law, and one of its central features is to impose duties on us to compensate one another for certain harmful accidents.[169] Similar compensatory intuitions can be found in commonsense morality as well.[170] Any appearance of inconsistency can nevertheless be dispelled by carefully distinguishing between the different kinds of rules that operate in these different areas of our normative practice. The discussions thus far—about the reactive attitudes and standard excusing conditions—apply to the criminal law and its near analogues in morality and law, wherever the rules in question lay direct claims on our conduct. It would, however, be a mistake to try to reduce too many of our normative practices to this simple model.[171] Hart, for example, argues forcefully that the law contains at least one other deeply important class of rules, which he calls "power-conferring rules."[172] These are not rules that primarily lay direct claims on our conduct, but rather ones that give us the power to vary our normative relations with one another and to create, modify or extinguish a number of important social relations and enterprises.[173] The law of contracts, marriage, wills and probate are familiar examples of this phenomenon, though commonsense morality contains numerous examples as well.[174] In my view, many of the rules of tort law represent yet another distinguishable class of rules, which might be called "liability-conferring rules." These are rules that do not, strictly speaking, tell us not to engage in any particular conduct, but they do tell us to be careful and to compensate others for certain losses occasioned by our accidents. [175]

---

[167] *See* FRANS DE WAAL, GOOD NATURED: THE ORIGINS OF RIGHT AND WRONG IN HUMANS AND OTHER ANIMALS 73-78 (1996).

[168] Holmes, *supra* note 166, at 238. George Fletcher has described this distinction between intentional and accidental as a universal feature of the criminal law. FLETCHER, *supra* note 38, at 111-30.

[169] Jules Coleman says that the "core of tort law is a certain practice of holding people liable for the wrongful losses their conduct has occasioned." *See* JULES L. COLEMAN, RISKS AND WRONGS 198 (1992).

[170] For the classic description of how principles of corrective justice animate aspects of our moral practice, see ARISTOTLE, NICOMACHEAN ETHICS bk.V.2, 4, at § 1130a-1131a, 1131B-1132b (Martin Ostwald trans., Prentice Hall 1999).

[171] For the classic argument for this proposition, see HART, *supra* note 5, at 27-28.

[172] *See id.* at 27-42, 44-49, 50-78, 91-99.

[173] *See id.* at 27.

[174] *See id.*

[175] This proposal is fully in line with Hart's basic insights. Hart himself says there is "some analogy" between the kinds of orders that most closely resemble criminal law and the law of torts, *id.* at 27, but he does not try to account for the distinctive normative force that tort rules have compared to either the criminal law or areas of the law like contract law. He also acknowledges that "[a] full detailed taxonomy of the varieties of law comprised in a modern legal system, free from the prejudice that all *must* be

Once these distinctions have been granted, the objection in question can be shown to be inappropriately focused. The perceived strength of the objection derives from the fact that many power- and liability-conferring rules appear to reflect resolutions to social contract problems, although the breach of these rules does not typically mark one out for exclusion in the sense needed to stabilize obligata. But power- and liability-conferring rules are importantly different from rules that lay primary claims on our conduct, and, when they resolve social contract problems, they do so in part because they are bound up with rules of this more basic kind.[176] These more basic rules typically require us either to perform certain acts or pay damages for the breach of contractual and related norms.[177] They typically require us to compensate others for harms we have occasioned in the case of accidents or unintentional torts, or—in the case of genuine strict liability—in some cases of harm where there is no fault at all.[178] The present account would thus predict that it is *these* rules that require stabilization, and this prediction is, in fact, borne out by the evidence.

For example, while expectation damages provide the most common remedy for contractual breaches in Anglo-American law, courts have sanctioning powers that they will employ to bring parties who intentionally refuse to pay civil damages awards into compliance.[179] Courts have similar powers to bring tortfeasors into compliance.[180] Corresponding to these legal facts are important facts about analogous moral situations: acts of apology and rectification for harms one has caused will tend to assuage moral aggression or hold it in abeyance, while intentional refusals to act in conciliatory manners will tend to provoke real ire.[181] In any event, punitive damages sometimes arise in tort law, and even—to a much lesser extent—in the law of contracts. They typically arise when, but only when, someone causes harm with the kind of mens

---

reducible to a single simple type, still remains to be accomplished," and he suggests his account of power-conferring rules is "only a beginning." *Id.* at 32. The account of liability-conferring rules proposed in the main text might thus usefully be viewed as part of a more developed taxonomy.

[176] *See, e.g.*, *id.* at 81 (explaining that such power-conferring rules are, in this sense, "parastic" on primary rules of conduct).

[177] Contract naturally give rise to performance obligations, but the standard remedy for contractual breach in Anglo-American is "expectation damages," which are meant to put the party in a position she would have been in if the contract were performed. *See* E. ALLAN FARNSWORTH ET AL., CONTRACTS 469 (6th ed. 2001). There has, however, been an expansion in the use of specific performance, and specific relief is the default in some civil law countries. *See id.* at 451-52. Punitive damages are, however, disfavored in both systems and are typically given only for failures to comply with court orders. *See id.* at 542-43.

[178] DAN B. DOBBS, THE LAW OF TORTS, § 1, at 2-3 (2001).

[179] Dan B. Dobbs, *Contempt of Court*, 56 CORN. L. REV. 183, 261 n.819 (1971); Diana Lowndes, *Authority of the Trial Judge*, 90 GEORGETOWN L.J. 1659, 1671 (1998).

[180] Lowndes, *supra* note 179, at 1671.

[181] For example, when parties continue to disagree rather than seeking conciliation, the empirical evidence suggests that this will tend to escalate conflicts. *See* DOUGLAS KNOLL, PEACEMAKING 304-07. Perceived moral imbalances like this can tend to persist if an apology is not made. *See id.* at 414-15.

rea ordinarily needed for criminal liability,[182] and this is precisely the shape that this account would predict our normative practices to take. Finally, further support for the present view could be found if the standard excuses were to creep back into our normative practices in response to these less common punitive reactions. This would appear to be the case: in both morality and law, we distinguish between the person who intentionally refuses to comply with a known duty to compensate and the person who has an excuse for failing to pay on time.[183]

Rather than undermining the present account, facts like these thus help show how it might unravel another puzzle about our normative practices. In particular, while standards of strict liability often crop up in tort law, there is a very strong presumption in the criminal law that some *mens rea* is required for criminal liability.[184] As one court has suggested:

> In the criminal arena, there is . . . a very strong presumption that some mental state is required for culpability. This requirement, which distinguishes those who perform acts knowingly, intentionally, or recklessly from those who perform them by accident or mistake, is "as universal and persistent in mature systems of law as belief in freedom of the human will and a consequent ability and duty of the normal individual to choose between good and evil."[185]

But as compelling as this principle is in the criminal context, there are large areas of tort

---

[182] W. PAGE KEETON ET AL., PROSSER AND KEETON ON TORTS, § 2, at 9 (5th ed. 1984).

[183] For example, we would be less likely to react harshly to someone who has sent a check in, but to the wrong address, or who has accidentally misspelled a name on a check and is willing to cure the defect. We would similarly be less likely to react harshly to someone who is forced not to pay at gunpoint or is prevented from paying by forces beyond her control. Modern law—especially in its highly bureaucratized form—may be less sensitive to these excusing conditions than commonsense morality, but there is an apparent sensitivity nonetheless. *See, e.g.*, JUDGE JAMES R. LAMBDEN ET AL., CALIFORNIA CIVIL PRACTICE: PROCEDURE § 30.79 (2001) (listing as defenses to a contempt sanction "inability to comply with judgment" and "good faith violation…[with present] willingness to comply").

[184] While there are some "strict liability" offenses in the criminal law, these are almost always offenses where either the punishment is minor, and more akin to a civil fine, or where there is a general mens rea requirement that is sufficient to put a person on notice that what one is is wrong and the strict liability attaches only to some more particular component of the action to which a heightened punishment is attached. Even cases of strict liability like these are few and far between in the criminal context. As H.L.A. Hart has suggested, "'strict liability' is generally viewed with great odium and admitted as an exception to the general rule, with the sense that an important principle has been sacrificed to secure a higher measure of conformity and conviction of offenders." H.L.A. Hart, *Prolegomenon to the Principles of Punishment*, *reprinted in* PUNISHMENT AND RESPONSIBILITY, *supra* note 6, at 20 (1968)

[185] United States v. Figueroa, 165 F.3d 111, 115 (2d Cir. 1998) (quoting Morrissette v. United States, 342 U.S. 246, 250 (1952)). H.L.A. Hart has similarly explained that "[i]t is characteristic of our own and all advanced legal systems that the individual's liability to punishment, at any rate for serious crimes carrying severe penalties, is made by law to depend, among other things, on certain mental conditions." *See* H.L.A. Hart, *Legal Responsibility and Excuses*, *supra* note 6, at 28.

law where the principle does not control.[186] This fact should be puzzling on its face, but—for the foregoing reasons—the present account would predict this distinction.

These facts thus provide another set of considerations that support the psychological views defended here. It would go beyond the scope of this Article to try to detail all of the ways the standard excuses operate in our moral and legal practices. The foregoing discussions should nevertheless clarify important ways in which they pervade these practices, and—so long as the proposition is understood to admit of further refinements like these—the pervasiveness of the standard excuses should be considered a part of the deep structure of morality and law.

F. ARGUMENT FROM AGENT-CENTEREDNESS

One objection that might be raised at this point is that the account is at odds with the "agent-centeredness" of commonsense morality and the law. Following standard convention, a requirement will be called "agent-centered" if, in at least some circumstances, it purports to give each person a different aim or goal, namely that *he* or *she* fulfill a given requirement even if by failing to do so that person could cause two or more others to fulfill the requirement in equally weighty circumstances.[187] A requirement will be called "agent-neutral" if, instead, it gives all people the same aims or goals.[188] A person who is under an agent-centered requirement not to break promises will, for example, sometimes be required not to break *her own* promises even if by doing so she might prevent two or more others from breaking theirs in equally weighty circumstances. A person who is under an agent-neutral requirement prohibiting promise-breaking would, by contrast, be required to minimize instances of promise-breaking—regardless of who the relevant promise breakers are. This agent-neutral requirement would not only allow but also require people to break their own promises if by doing so they could prevent two or more others from breaking promises in equally weighty circumstances.

As the last example illustrates, many imperatives can be stated in ways that leave them ambiguous as to whether they are agent-centered or agent-neutral. Properly construed, commonsense morality and the law are, however, replete with agent-centered restrictions.[189] This fact might be thought to pose a problem for the present account for a simple reason: some social contract problems can seemingly be resolved by

---

[186] *See, e.g.*, DOBBS, *supra* note 178, § 392, at 1097-98 (strict liability in workers' compensation); *id.* § 346, at 950-52 (strict liability for abnormally dangerous activity); *id.* § 334, at 905-06 (employer liability for torts of employees).

[187] *See, e.g.*, PARFIT, *supra* note 102, at 27, 54-55; SAMUEL SCHEFFLER, THE REJECTION OF CONSEQUENTIALISM 80 (1982) (explaining that a theory will contain an agent-centered restriction if "there is some restriction S, such that it is at least sometimes impermissible to violate S in circumstances where doing so would prevent a still greater number of equally weighty violations of S, and would have no other morally relevant consequences"); ELIZABETH ANDERSON, VALUE IN ETHICS AND ECONOMICS 73 (1993). It is widely accepted that common sense morality is pervaded by agent-centered restrictions, though some have argued that the authority of such restrictions is problematic.

[188] *See, e.g.*, PARFIT, *supra* note 102, at 27, 54-55.

[189] *See, e.g.*, Stephen Darwall, *Introduction*, *in* DEONTOLOGY 1-7 (Stephen Darwall ed., 2003).

adopting broad agent-neutral standards, which give each member of the relevant social contract a shared aim or goal. "Act for the common good" would be one such plausible standard, and a number of economists, including Richard Posner, have argued that adopting a fundamentally contractarian decision procedure will, in fact, result in a standard that requires us to maximize efficiency or pareto-optimal states of affairs.[190] This standard is agent-neutral; yet commonsense morality and the law have a different face to them, and their standards typically permit us to refrain from some actions that would otherwise be efficient or tend to the common good.[191] Indeed, this is a central feature of their binding nature, a fact that David Hume famously observed and illustrated with the following example: the miser who owns a bit of property to which he attaches no real value is still typically viewed as morally and legally entitled to its recovery, even if someone who has taken it would enjoy it more.[192] More generally, the fact that a given piece of property is *mine* or *yours*, that a given promise is *mine* or *yours*, and the like, can make an important moral and legal difference—one which cannot be accounted for in purely agent-neutral terms. If morality and law engage attitudes that naturally function to allow us to resolve social contract problems, then why do they purport to present us with so many relatively simple agent-centered restrictions rather than a broad agent-neutral standard that in fact resolves social contract problems?

This challenge is certainly not decisive. Evolutionary dynamics do not typically produce optimal results, and many of the simpler rules of action that we find in morality and law do in fact resolve social contract problems. Hence, while broader, agent-neutral standards like "act for the common good" might appear optimal in theory, a supporter of the views developed here might try to account for the agent-centered features of morality and law as arising from evolutionary suboptimalities. This line of response would, however, be unsatisfying—or at least incomplete—in light of the pervasiveness of simple, agent-centered restrictions in morality and law.

A more penetrating response would show that even perfect capacities to resolve social contract problems would leave us perceiving ourselves under a multiplicity of simpler rules for action that are agent-centered in form. The literature on rule utilitarianism helps with one part of this project. It suggests why, even if our shared aim or goal were to maximize utility impartially assessed, sharing a maxim with that content

---

[190] *See, e.g.*, Richard A. Posner, *The Ethical and Political Basis of the Efficiency Norm in Common Law Adjudication*, 8 HOFSTRA L. REV. 487, 488-502 (1980); John C. Harsanyi, *Bayesian Decision Theory and Utilitarian Ethics*, 68 AM. ECON. REV. 223-28 (1978); Hal R. Varian, *Distributive Justice, Welfare Economics, and the Theory of Fairness*, 4 PHIL. & PUB. AFF. 223, 228-29, 240 (1973); Kenneth Arrow, *Some Ordinalist-Utilitarian Notes on Rawls's* Theory of Justice, 70 J. PHIL. 245-50 (1973).

For important philosophical criticisms of these arguments, see, e.g., RAWLS, *supra* note 65, at 14, 118-92; T.M. Scanlon, *Contractualism and Utilitarianism*, *in* UTILITARIANISM AND BEYOND, *reprinted in* CONTRACTARIANISM/CONTRACTUALISM, *supra* note 67, at 219, 235-43; Jules L. Coleman, *Efficiency, Exchange, and Auction: Philosophic Aspects of the Economic Approach to Law*, 68 CAL. L. REV. 221, 237-47 (1980); Jules L. Coleman, *Efficiency, Utility, and Wealth Maximization*, 8 HOFSTRA L. REV. 509, 525 (1980).

[191] For a description of these and a number of related intuitive problems for act consequentialists, see Richard B. Brandt, *Toward a Credible Form of Utilitarianism*, *reprinted in* CONSEQUENTIALISM 207, 209 (Stephen Darwall ed., 2003).

[192] *See* HUME, *supra* note 104, at 502.

would likely be less optimal than internalizing simpler rules that have a more familiar moral and legal look.[193] Richard Brandt has collected the most commonly cited examples of this phenomenon. First, it can be difficult to apply the utilitarian calculus, thus making it likely that simpler standards for action, which take into account our limited intelligence and other cognitive weaknesses, will maximize utility.[194] Second, people tend to rationalize in their own favor. Hence, concrete standards that allow for less judgment in application can better serve the common good than broad standards, which are ambiguous as to what they rule out.[195] Third, we must often act quickly, and without time for adequate deliberation. Deliberation can also be costly. Hence, relatively simple rules—which give us specific directions in recurrent and readily-identifiable situations—can sometimes conduce to the common good better than a general utilitarian standard. Fourth, the collective following of an act utilitarian standard can be self-defeating, and we sometimes need to coordinate to achieve the common good. At times, simpler coordination rules are thus our most direct route to the common good.[196] Finally, it is often complained that broad utilitarian standards impose oppressive demands on us and leave too little room for individual freedom to pursue our own ends.[197] A certain sphere of autonomous choice may be needed to allow us to achieve personal happiness, an important ingredient of the common good.[198]

The psychological arguments presented thus far suggest a further, underappreciated reason why, on the present assumptions, we might expect our moral and legal codes to contain a multiplicity of simpler rules for action. Proponents of broad consequentialist maxims like "act for the common good" typically view the question of how to act and how to react to deviations from moral and legal norms as separable questions.[199] There are, however, familiar costs associated with punishing deviations from norms, and it is certainly not the case that punishing every action that fails to live up to an act utilitarian standard would maximize utility.[200] In proposing broad agent-neutral

---

[193] *See, e.g.*, BRANDT, *supra* note 6, at 111-36; Peter Railton, *How Thinking about Character and Utilitarianism Might Lead to Rethinking the Character of Utilitarianism*, 13 MIDWEST STUD. PHIL. 398 (1988), *reprinted in* RAILTON, *supra* note 79, at 226-28; Brandt, *supra* note 191, at 209, 213-25; John Rawls, *Two Concepts of Rules*, 64 PHIL. REV. 9-13 (1955); ROSS, *supra* note 3, at 38-39; DAVID LYONS, FORMS AND LIMITS OF UTILITARIANISM 10-12 (1965).

[194] *See* BRANDT, *supra* note 5, at 273.

[195] *See id.* at 232, 274.

[196] *See id.* at 274.

[197] *See id.* at 276-77.

[198] Although rule utilitarians generally adduce considerations like these to try to justify adopting a rule utilitarian standard of the right rather than to explain the look of our moral and legal practices, all of these considerations can be imported into the present explanatory context.

[199] For example, act utilitarians typically define the "right" act as that act that maximizes or conduces to the "good," and then typically view the question of what punishment is right or justified as a separable instance of this question, relating to whether a particular act of punishment will maximize or conduce to the good. *See, e.g*, C.L. TEN, CRIME, GUILT AND PUNISHMENT 7 (1987). Brandt also assumes a certain amount of separability in his discussions of human psychology. *See* BRANDT, *supra* note 5, at 174, 175, 286, 288, 291, 293.

[200] Indeed, Brandt has argued for what he calls the "revolutionary implication" of utilitarianism: "[I]t is clear that there is a prima facie case against the moral code prohibiting anything—which may be quite

standards for action, consequentialists like Derek Parfit are thus forced to try to define normative categories like those of "blameless wrongdoing."[201] The arguments in this Article suggest, however, that obligata are portfolios of primary motive and reactive attitude, which come together as part of a distinctive syndrome. Indeed, the suggestion is that these elements must come together for obligata to arise and persist in nature—and, moreover, that we need obligata to identify and respond to moral and legal obligations. It is therefore unclear whether the identification of categories like "blameless wrongdoing" can hold our attention in the right way to reflect practicable normative proposals. At the very least, it will continue to be a centrally important—and perhaps the most important and pressing—normative question to determine what is right, and what wrong, where these terms are used in their familiar senses, as entailing both that we have exclusionary reasons to act and reasons to react to deviations in the ways discussed here.[202] But this means refusing to separate these questions in central areas of normative inquiry, and instead focusing on which binding moral or legal rules would conduce to the common good. There is a large literature in the rule utilitarian tradition the answer to this question is a multiplicity of standards for action that are very much like the ones we commonly find in morality and law.[203]

These considerations provide a first step to answering the puzzle that began this section, but they do not go all the way. They suggest why, on the present assumptions, we might expect morality and law to contain a multiplicity of relatively simple and familiar rules for action, rather than broad consequentialist mandates. But they do not explain why these simpler rules would appear in an agent-centered rather than agent-neutral form. The proposed explanation for simple rules is that they sometimes conduce to certain goals we might agree to share, in order to resolve various social contract problems, and sometimes do so better than direct mandates to seek those goals. If acting in accordance with simple rules like these will have these consequences, however, then why would our moral and legal practices not allow us to violate these rules if we could thereby prevent two or more others from engaging in equally weighty violations? Legal and moral obligations often prohibit us from acting in this way, and this is the deeper puzzle that their agent-centeredness poses for the present account.[204]

---

surprising. For what someone wants to do there is (at least normally) some benefit in permitting: he will enjoy doing it, and feel frustrated in being prevented on grounds of conscience. If something is to be prohibited or enjoined, a case must be made out for the long-range benefit of restricting the freedom of individuals, making them feel guilty, and utilizing the teaching resources of the community." BRANDT, *supra* note 5, at 293.

[201] *See* PARFIT, *supra* note 102, at 31-35.

[202] This is presumably why we so naturally think the fact that someone has done wrong is a reason to criticize that person's actions. *See, e.g.*, MILL, *supra* note 143, at 193.

[203] For a classic argument of this form by a rule utilitarian, see Brandt, *supra* note 191, at 207, 209, 213-25; *see also* SCHEFFLER, *supra* note 187, at 112. For a deontologist's description of rule utilitarianism that is sympathetic to this quality of the theory, see Darwall, *supra* note 102, at 119.

[204] The problem in question here is distinct from the more well-known puzzle of how—if at all—agent-centered restrictions might be justified. This latter question is one of the central questions in normative ethics. *See, e.g.*, Darwall, *supra* note 102, at 112-38. The problem here is, however, purely explanatory, and relates only to the plausibility that the explanatory framework presented would plausibly result in psychological attitudes that fixate us on rules that are agent-centered in form.

The appearance that there is an inconsistency here might nevertheless be dispelled by attending to the distinctive epistemological difficulties we typically face when applying agent-neutral as opposed to agent-centered standards to the facts. On the present assumptions, these difficulties should affect not only our first personal deliberation when deciding what to do but also our second personal deliberation when deciding how to react to one anothers' apparent deviations.[205] Moreover—and this is *key*—it will typically be much easier to identify breaches of a norm in its agent-centered rather than its agent-neutral form. This is because there is only one relevant causal link between an agent and an action required of her by an agent-centered standard, whereas the causal links between an agent and the sum total of acts by anyone of a particular type can be very long, very complex, and very difficult—if not impossible—to ascertain. To illustrate with the familiar example that began this section, it will typically be much easier to determine whether a person has broken her own promise than whether a person has acted in a way that causes fewer promises to be kept. This point is, moreover, generalizable, either fully or to a very significant degree. Rather than providing a counterexample to the present psychological account, the deeply agent-centered nature of morality and law is thus a feature that contributes to their stability on the present assumptions.

These last arguments will, moreover, apply with even greater strength if—as is more plausible—evolutionary forces have left us with merely suboptimal psychological capacities to resolve social contract problems, which incline us to share rules that conduce to the common good rather than maximizing it. These arguments will similarly apply if—as a number of people have argued very plausibly—the resolution of certain social contract problems is not one broad consequentialist standard but rather a number of simpler shared standards for action that are more familiar from morality and law.

These considerations thus allow for the final promised refinement of our definition of obligata—as set forth in element (vii). For while the initial definition of obligata began with a seemingly epistemic notion of suppositions—which were merely suppositions *that* all (or a significant majority of) others are relevantly committed—this more complex functional state is better characterized as involving suppositions *of* each relevant individual that he or she will fulfill a standard, which normative supposition is exhibited in us in tendencies to hold others accountable, to make claims on their conduct, to criticize deviations, to resent what they have done to us, and the like.[206] To serve their function well, these attitudes will also need to focus us on relatively simple rules for action that are agent-centered in form, and these features of morality and law should thus also be understood as part of their deep structure.

---

[205] The second personal standpoint is the standpoint we use when we address one another with claims and grievances. This is the standpoint that Stephen Darwall has recently argued to play a central role in our normative thought and practice. *See* STEPHEN DARWALL, MORALITY AND THE SECOND PERSON STANDPOINT (forthcoming, on file with author).

[206] Though there may be differences at the margin, this dimension of our moral and legal practices is—on my view—intimately related to what Stephen Darwall has called the "second-personal standpoint." *See id.*

G. SPECIAL NORMATIVE TERMINOLOGY

Consider an important admission that Joseph Raz has made. Although a staunch legal positivist, he concedes that: "One of the main stumbling blocks for legal positivists has been the use of normative language, i.e., the very same terminology which is used in moral discourse, in legal discourse. The fact that the law is described and analysed in terms of duties, obligations, rights and wrongs, etc., has long been regarded by many as supporting the claim of the natural lawyer that law is inescapably moral."[207] One of the central problems in moral theory is, in fact, to determine what we might even mean when we use this special terminology and make claims about things like "obligations" or that a given action is "right" or "wrong."[208] To many, it seems implausible that we could be referring to objective properties with this language, at least if the properties are to have the prescriptivity or intrinsic motivational force that obligations purportedly have.[209] For how could properties of this kind exist in the natural world?[210] Moreover, even if such properties were to exist—say, in some non-natural realm—it should seem equally puzzling how we might have epistemological access to this realm, or how these non-natural properties might causally interact with us so as to produce reliable knowledge.[211]

One might think the law is importantly different in this regard because the law arises wholly from social conventions. There are now a number of plausible accounts of social facts in terms of social conventions, which render these facts wholly unmysterious and consistent with a broadly naturalistic worldview.[212] One might thus try to account for legal facts—*e.g.*, facts about whether we have certain "legal obligations," or whether certain actions are "legally permitted" or "legally required"—on this same basic model. Such an account would not only vindicate our common intuition that there can be facts of the matter about what the law requires but would do so without requiring us to posit any strange, non-natural moral properties or implausible epistemological

---

[207] RAZ, *supra* note 1, at vii; *see also* HART, *supra* note 5, at 85-86.

[208] Questions about the meaning or status of moral (or other normative) terms are typically referred to as "meta-ethical" questions to distinguish them from questions about what morality (or other normative areas like the law) require in a given set of circumstances. *See generally* Stephen Darwall et al., *Toward a* Fin de Siècle *Ethics*, 101 PHIL. REV. 115 (1992). To know what morality and law require is not yet to know what it *means* for morality or law to require something.

[209] *See, e.g.*, J.L. MACKIE, INVENTING RIGHT AND WRONG 38-40 (1977).

[210] For classic exposition, see *id.* at 38-39. Mackie says that such properties would be "queer properties," unlike anything else we are familiar with in the natural world. *Id.* at 38.

[211] For classic exposition, see *id.* at 38, 41. Another puzzle arises from the fact that moral and legal judgments typically supervene on natural facts, and we often explain our judgments by citing those facts. But how could natural facts *explain* facts in some other non-natural realm without causally interacting with them? *See id.* at 41.

[212] For an exemplary and representative account of this kind, see MARGARET GILBERT, ON SOCIAL FACTS (1989).

capacities. The modern tradition of legal positivism can, in fact, be viewed as in large part an attempt to make good on this basic promise.[213]

It has nevertheless proven notoriously difficult for legal positivists to account for certain facts about adjudication in these terms. Our practices of adjudication appear to allow for fundamental disagreement over the criteria that allow us to identify the law, but disagreement of this particular kind is inconsistent with the claim that we identify the law by using a social convention in the sense employed by most legal positivists. [214] Whether legal positivists can plausibly accommodate these features of adjudication into their accounts of the law—or can otherwise establish that these facts have no genuine relevance to the inquiry—is an outstanding question for legal theory.[215] Even if the legal positivists were to succeed, however, it would remain puzzling why the law—now viewed as an institution that is distinguishable from morality in part because we use a social convention to identify the law—would employ the same special vocabulary as morality. How might one account for whatever is the same about the meanings of these special terms as they arise in morality and law?

One promising approach to trying to account for our moral language is to begin with the question of what psychological state we are in when we sincerely *believe* we are under under a moral obligation, and then to account for the meaning of our special moral terminology as in some sense expressive of these psychological states.[216] If the sincere expression of terms like "ought," "obligation," "duty," "right," and the like were expressive of obligata, then—for reasons already discussed—this fact would explain why, in both morality and law, the thought that one has an "obligation" is typically taken to entail that one has a sufficient and agent-centered reason to fulfill that obligation. This proposal would also explain why we taken reasons arising from obligations as capable of excluding others that arise from antecedent desire or personal interest, and why we take failure to fulfill such obligations as warranting certain forms of reaction and/or punishment or coercion for non-compliance.[217] The structural complexity of obligata would—on this view—distinguish the thought that something is an obligation from the

---

[213] *See, e.g.*, HART, *supra* note 5, at 79-123, 254-59 (accounting for rule of recognition in terms of social conventions); Jules L. Coleman & Brian Leiter, *Legal Positivism, in* A COMPANION TO PHILOSOPHY OF LAW AND LEGAL THEORY 241, 243 (Dennis Patterson ed., 1996) (stating that inclusive legal positivists assert that "what norms count as legal norms in any particular society is fundamentally a matter of social conventions"); Jules Coleman, Legal Positivism Since H.L.A. Hart (unpublished manuscript, on file with author) (noting that this is the general feature that unites modern positivist theories).

[214] For an exemplary discussion of these problems and of the current state of the debate by a legal positivist, see Coleman, *supra* note 213.

[215] *See id.*

[216] This basic pattern of analysis is called an "expressivist" analysis. Allan Gibbard has, for example, proposed an expressivist analysis of rationality as follows: "To call something rational is not to attribute some particular property to that thing—not even the property of being permitted by accepted norms. . . .We explain the term by saying what state of mind it expresses." GIBBARD, *supra* note 111, at 7-8, *see also id.* at 45-48.

[217] *See, e.g.*, MILL, *supra* note 143, at 193 ("We do not call anything wrong unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience.").

mere thought that we have a reason to act that arises from various things we desire or conduce to personal interest. There may, of course, be a further fact—which legal positivists have perhaps rightly latched onto—concerning the distinctive role that social conventions play in identifying valid legal but not moral requirements, but none of this would undermine the power of the present account to illuminate why morality and law share the same special vocabulary.

The special normative vocabulary that morality and law share arises in a number of different contexts. Sometimes we use it to discuss what the law is or what morality requires, to tell stories with various morals, or to gossip about one anothers' transgressions.[218] Another important use of this language is, however, to address one another second-personally with claims or grievances.[219] When we do this, we are charging one another with having done something wrong or against the law. But to do something "wrong" or "against the law" in this sense, one must not only perform an act that is "wrong" or "against the law" as these terms appear less charged contexts. One must also do so in a way that reveals we are improperly motivated by morality or law. As earlier discussions have suggested, our second personal attitudes towards one another should have precisely these features: obligata consist in part of reactive attitudes that fund our second-personal reactions and claims against one another, and some responsiveness to the standard excuses should be an expected part of the deep structure of these attitudes. The fact morality and law share the additional special normative vocabulary of "standing," "claim," "charge," "complaint," "grievance," "excuse," "justification" and the like would thus be well explained by citing structural features of the attitudes that give morality and law their common life. The special normative vocabulary that morality and law share should thus—on the present account—be viewed as yet another one of their deep structural features.

Before continuing, an important clarification is in order. The general tack of trying to account for aspects of the meanings of our moral terms as expressive of various psychological attitudes has historically been associated with so-called "non-cognitive" accounts of this language, which deny that moral judgments can be true or false.[220] More recent theorists like Allan Gibbard and Simon Blackburn have, however, developed expressivist accounts of moral language that accommodate many of its objectivist features, including the way we meaningfully disagree, the way we embed moral language into conditionals, the way we draw inferences from moral premises to moral conclusions, and the fact that our moral language supervenes on natural facts.[221] These so-called "quasi-realists" have adopted a "minimalist" account of truth,[222] and have

---

[218] *See, e.g.*, GIBBARD, *supra* note 111, at 71-74; JOHN SABINI & MAURY SILVER, MORALITIES OF EVERYDAY LIFE (1982).

[219] *See, e.g.*, DARWALL, *supra* note 205; HART, *supra* note 5, at 90.

[220] *See, e.g.*, Justin D'Arms & Daniel Jacobson, *Sentiment and Value*, 110 ETHICS 722, 730 (2000) ("Traditionally, noncognitivists denied that value judgments are apt for truth . . . .").

[221] For the most important developments of this line of thought, see ALLAN GIBBARD, THINKING HOW TO LIVE 44-111 (2003); SIMON BLACKBURN, ESSAYS IN QUASI-REALISM (1993).

[222] Minimalists about truth reject the notion that truth is a natural property of sentences, and instead try to explain what we mean by this predicate in terms of a set of sentences of the form "The sentence 'P' is true

suggested that there may be no property of truth that our descriptive judgments can have that our moral judgments must lack.[223] Quasi-realists nevertheless still deny that there is any objective property of rightness that might explain our moral judgments,[224] and it is therefore important to recognize that this Article neither endorses nor rejects such claims. As with the case of concepts like "desirable" and "good," there is still room, in my view, for reasonable disagreement on this issue. The question whether there are objective properties that might explain our moral or legal judgments will likely depend on two things: (i) whether we can settle on a normatively satisfying account of what is required of us by morality or law; and (ii) whether the correct evolutionary explanation of our capacities for normative judgment shows that they function to identify the natural facts upon which these requirements supervene.[225] It would go beyond the scope of this Article to try to answer these questions here, but it is important to recognize that these questions have been left open.

H. ESSENTIAL CONTESTABILITY AND CONTENT

Up until this point, everything that has been said would, in principle, be consistent with the view that natural selection has endowed us with certain natural and inescapable views on right and wrong, which allow us to resolve discrete classes of social contract problems that we faced recurrently in ancestral circumstances. A number of people working in evolutionary game theory, such as Brian Skyrms and Robert Sugden, have, in fact, recently begun to suggest specific game theoretic explanations for the emergence of particular norms, such as those of private property.[226] There is also some evidence that certain normative commitments appear broadly, and cross-culturally, in human life. Hart, for example, once suggested that while the moral and legal codes in

---

if and only if P." *See, e.g.*, Alfred Tarski, *The Semantic Conception of Truth and the Foundations of Semantics*, 4 PHIL. & PHENOMENOLOGICAL RESEARCH 341, 343 (1944).

[223] *See, e.g.*, GIBBARD, *supra* note 221, at x ("Does this mean there are no facts of what I ought to do, no truth and falsehoods? Previously I thought so, but other philosophers challenged me to say what this denial could mean. In this book, I withdraw the denial and turn noncommittal. In one sense, there clearly are 'facts' of what a person ought to do, and in a sense of the word 'true' there is a truth of the matter. That's a minimalist sense, in which 'it's *true* that pain is to be avoided' just amounts to saying that pain is to be avoided—and likewise for 'it's a fact that.'").

[224] *See, e.g.*, GIBBARD, *supra* note 221, at 251-67.

[225] For a helpful description of how such an explanation would have to work if one were to maintain commitment to a minimalist conception of truth, see *id.* Gibbard does not believe that such an explanation will be forthcoming for our practical judgments. *See id.* at 267 ("I have been baffled in seeing how any such story could be told."). This Article remains agnostic on that front.

[226] *See* SKYRMS, *supra* note 113, at 76-79; Sugden, *supra* note 88 (developing an evolutionary game theoretic explanation of the emergence of de facto property rights). Sugden's analysis importantly presumes that we have a propensity to form normative expectations, which consist not only in expectations of conformity to shared behavioral regularities but some motivation to meet other peoples' expectations, thus marking an important difference from standard game theoretic assumptions about human motivation.

particular societies can vary a great deal, there is still a set of minimal normative commitments that all societies share, which he identifies as containing norms for the "protection of persons, property and promises."[227] Hart says that "[s]uch rules do in fact constitute a common element in the law and conventional morality of all societies which have progressed to the point where these are distinguished as different forms of social control."[228]

For a number of reasons, however, it would appear to be a mistake to view the contents of our sense of obligation as hard-wired at the level of content. As an initial matter, it must be conceded that for any given candidate universal norm, there is genuine controversy over its claim to cultural universality.[229] To the extent that cultures converge on specific kinds of norms, a close examination often reveals that there are numerous and noteworthy distinctions between the precise content and status of these norms in different cultural settings. For example, for those who think that practices of contract and promise-keeping are absolutely central to human social life, Marcel Mauss's classic work *The Gift* is instructive because it indicates that in many so-called "primitive" societies, goods have rarely been exchanged by means of explicit contracts or other reciprocal promises in market transactions.[230] Goods have instead typically been exchanged by means of intricate systems of reciprocal gift-giving, in which the owners of property are viewed as obliged to give gifts to those with whom they associate, thereby generating a shared sense of obligation to reciprocate.[231] The conditions under which violence is tolerated (and sometimes even expected) is also something that can differ quite vastly from culture to culture—as anthropological studies of aggression have shown.[232] Sometimes, finally, the surface universality of a candidate norm lies at a level of generality that is one step removed from its precise content. If, for example, property norms are defined in the paradigm case as norms that give individuals or collective

---

[227] HART, *supra* note 5, at 199. Hart is careful to point out that he does not view the binding force of such norms as rationally necessary, and instead views them as necessary given a number of contingent but robust facts about human beings, human nature and the world we find ourselves in. These include the facts that (i) we generally have the aim of survival, (ii) we are physically vulnerable, (iii) none of us is so much more physically powerful than others that we can dominate or subdue all others, without some cooperation from others, for more than a short period of time, and (iv) we have limited altruism, limited resources and limited understanding and strength of will. *Id.* at 193-98.

[228] *Id*. at 193.

[229] For example, it is quite common in the anthropological literature to insist—rightly or wrongly—that "the claims of ethical propositions are relative to the moral standards of the cultures in which they are embedded." Melford E. Spiro, *Cultural Relativism and the Future of Anthropology*, 1 CULTURAL ANTHROPOLOGY 259, 260-61 (1986).

[230] *See* MARCEL MAUSS, THE GIFT: THE FORM AND REASON FOR EXCHANGE IN ARCHAIC SOCIETIES 3-46 (W.D. Halls trans., W. W. Norton & Co. 1990) (1950).

[231] *See id.*

[232] For a good discussion of cultural variability in aggressiveness leading to homicide, which tries to square the record with an adaptationist account of aggression, see DALY & WILSON, *supra* note 150, at 275-97. Daly and Wilson also discuss cultural variations in the legitimacy of violence, and some of the anthropological and ethnographic evidence relating to this variation. *See id.*

bodies various bundles of rights to use, to exclude others from the use, and/or to transfer certain goods,[233] then every culture probably has property norms of some sort, which more or less closely resemble this paradigm, but cultures differ so radically over what goods can be acquired, what precise bundles of rights come with the acquisition, and how specific property rights are created, varied, limited, extinguished and apportioned among individuals or groups that the search for universal property norms would appear to be in vain.[234]

Of course, the fact that there are controversies over the universality of any given normative commitment does not mean that there is no fact of the matter, and one might try to resolve some of this initial class of problems in favor of a narrower set of core normative commitments that are in fact universal in the relevant sense. Still, even if there is some such set, the ethnographic record displays such a surprisingly wide array of norms in our moral and legal codes[235] that the set of remnant commonalities—if indeed there are any—clearly comprises a very small, and arguably negligible, portion of our normative practices.

There is, moreover, another important reason to think that even universal human agreement in views about the right would not render these views inescapable, one which relates more centrally to the logic and nature of our normative practices. When we accept propositions about what is right, we sometimes express our acceptance by making judgments that prescribe certain courses of action or conduct. It is a well-known fact, however—at least since G.E. Moore framed his famous "open question" argument[236]—that when we make such judgments, we can always meaningfully ask whether the actions or courses of conduct that we disapprove of are, genuinely, wrong.[237] Related to this is another well-known fact: for any moral judgment that we make, it is always possible for others coherently and meaningfully to disagree with the assessment.[238] Where there is the possibility of disagreement, it is plausible to think that there is the possibility of

---

[233] JOHN AUSTIN, LECTURES ON JURISPRUDENCE: OR THE PHILOSOPHY OF POSITIVE LAW § 1000 (4th ed. 2002) (1875).

[234] There may be a number of basic normative categories that we are inclined to find salient, but that can be worked out in a number of different ways in different cultural settings. Alan Fiske has, for example, proposed that we have four basic modes of social relations—all of which would resolve social contract problems and so would be consistent with the more general analysis offered here. *See* FISKE, *supra* note 120. Categories like those of property may also have a distinctive structure worth elaborating. Nothing in this Article should be viewed as either endorsing or rejecting more specific proposals like this.

[235] *See, e.g.*, Spiro, *supra* note 229, at 262; FLETCHER, *supra* note 38, at 4.

[236] MOORE, *supra* note 3, at 6-20

[237] *See supra* note 3.

[238] *See, e.g.*, Gallie, *supra* note 2, at 167-78. Allan Gibbard has recently proposed that it is ultimately this feature of moral discourse that best explains the continuing importance of Moore's investigations. *See* Allan Gibbard, The Reasons of a Living Being, Presidential Address Before the Central Division, *in* 76 PROC. & ADDRESSES AM. PHIL. ASSOC. 49-60 (2002), *http://www.lsa.umich.edu/philosophy/Gibbard.htm.* Gibbard extends this claim to normative judgment more generally. *See id.*

persuasion or conversion.[239] Indeed, modern psychological research suggests that we are sometimes subject to persuasion or conversion of this kind in contexts of discussion with perceived peers, even sometimes when those changes are not traceable to reasons that we antecedently accept.[240] Our moral and legal judgments are thus apparently essentially contestable, even if widely or universally shared, and this essential contestability is an ineliminable feature of the attitudes we express with these judgments. Moreover, even if there were universal agreement that a given action were morally or legally required, this fact could alone not establish that the action was required on pain of regress, because it is at best an open question whether the fact of convergence bears on any particular moral or legal question.

One could, on the other hand, explain the full range of facts under discussion by positing that we have obligata that allow us to resolve social contract problems in a flexible manner. This particular kind of flexibility would allow for the cultural variation we see in moral and legal views.[241] This explanation would, moreover, explain equally well why our ethnographic record exhibits certain recurrent and common normative commitments. [242] There are a number of familiar problems that we face in human life that are so common and widespread, and that appear in so many different real life circumstances, that a flexible capacity to resolve social contract problems should be expected to produce some common commitments.[243] The purportedly universal and necessary commitments that theorists often point to—like those concerning promising, norms against wanton violence, and the like—typically have this quality to them.

To flesh out this form of explanation, one would need to identify plausible psychologically mechanisms that might allow us to adapt our moral and legal codes to changing social contract problems. Given the nature of obligata, any such mechanism would need to meet a number of constraints. To see why this is so, notice, first, that our moral psychologies will generate an important biological need for coordination over the perceived content of the right, given the present assumptions. This need derives from a fact that has already been discussed: for obligata to resolve social contract problems and allow us to reap the benefits of this cooperation, they must be bound up with second

---

[239] Justin D'Arms and Daniel Jacobson have, for example, parsed the essential contestability of our normative concepts as leaving us with the following moral: "[T]he very notion of essential contestability seems to be that nothing settles such questions; rather, the function of these concepts requires that their application remain open to normative influence—the giving and taking of reasons." D'Arms & Jacobson, *supra* note 220, at 738.

[240] *See* Jonathan Haidt, *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, 108 PSYCHOL. REV. 814-34 (2001).

[241] *See supra* note 235.

[242] This flexibility in nevertheless perfectly consistent with the deep structural features of obligata that have been discussed thus far. This flexibility is also consistent with the fact that legal norms may have to meet certain basic criteria to be law-like, as Lon Fuller has maintained. *See* LON FULLER, THE MORALITY OF LAW 33-41 (1964). Fuller is clear that he does not believe the criteria he identifies (which he calls the "internal morality of law") entail anything about the substantive moral content of the law. *Id.* at 152-55.

[243] HART, *supra* note 5, at 90.

order attitudes that function to identify and exclude non-cooperators from the benefits of these cooperative enterprises.[244] Secondary attitudes of this kind are, however, inherently non-cooperative—and, in fact, often strongly so. Hence, they should involve tendencies to action that would otherwise be ruled out by the code itself, except in cases where a prior breach warrants the reaction. Where the contents of the obligata in a group are not coordinated, this means—somewhat paradoxically—that the very moral psychologies that allow us to resolve social contract problems, and engage in cooperative social living, create the possibility of escalating conflict.[245] These cycles can occur when actions that are perceived as a wrong by others, and hence as warranting certain critical or punitive reactions, produce actions that are, in turn, perceived as a wrong that requires a further righting—and so on down the line.[246]

This need for coordination will, moreover, take on a very particular quality to the extent that our normative psychologies give us a flexible capacity to resolve changing social contract problems. As already noted, social contract problems are a species of commitment problem,[247] and, hence, they require for their resolution motivations that cannot be abandoned on just any ground or in just any circumstance. When adapting our views, maintaining an appropriate modicum of coordination will also be essential. Hence, any capacity to adapt our views should be ones that allow us to do so while maintaining both (i) an appropriate modicum of intrapersonal commitment to the relevant standards of a group and (ii) a similar modicum of interpersonal coordination over their content.[248]

Can we identify psychosocial processes that meet these criteria? Allan Gibbard has recently described our capacities to engage in "normative discussion" as having just these features. Gibbard argues that the attitudes that are principally involved in familiar cases of moral judgment are ones that not only motivate action and certain moral emotions, like guilt and impartial anger, but also manifest themselves in tendencies to expression and avowal in the special terminology familiar from moral discourse in contexts of what he calls unconstrained "normative discussion."[249] Gibbard defines "normative discussion" as including not only substantive moral debate but also a broad range of other important phenomena that are common in human life—things like gossip,

---

[244] *See* Section D, *supra*.

[245] Hart cites this as one of the primary problems that moves us from a pre-legal into a legal state. HART, *supra* note 5, at 93-94.

[246] For a particularly vivid description of how such conflicts can escalate, see ROBERT FRANK, PASSIONS WITHIN REASON 1-4 (1988); *see also* HART, *supra* note 5, at 93 ("It is obvious that the waste of time involved in the group's unorganized efforts to catch and punish offenders, and the smouldering vendettas which may result from self help in the absence of an official monopoly of 'sanctions,' may be serious.").

[247] See *supra* notes 90-93 and accompanying text.

[248] We should, in other words, expect to find less genuinely action-guiding modification to occur via purely private reflection and/or in non-social contexts and most such modifications to occur through social processes, which allow the members of a group to shift their views together.

[249] GIBBARD, *supra* note 111, at 64-80.

the discussion of stories and movies, our interest in "I was like . . . ; He was like" conversations—and the like.[250] Canvassing a number of broad theoretical and evolutionary considerations, Gibbard argues that the biological function of normative discussion is to coordinate our normative views.[251] He proposes that this coordination is fostered by tendencies toward mutual influence and by mutual demands for consistency, along with corresponding inclinations to subject our moral views to norms of consistency.[252] Although Gibbard's account of what attitudes are coordinated in normative discussion differs in some important details from the psychological claims defended here, there is also a great amount of resonance. Gibbard's arguments about normative discussion will apply with equal force even if—as this Article claims—normative discussion were to coordinate obligata rather than the closely related attitudes that Gibbard has identified.

There is, moreover, now further empirical support for the core aspects of normative discussion under examination. Jonathan Haidt has recently collected psychological research suggesting that a significant number of our moral convictions are held as relatively automatic judgments attuned to particular cases, which he calls "intuitions."[253] When we take certain things to be morally wrong, we often rely on these perceptions, and, although we take ourselves to be justified in our views and can sometimes cite reasons for our views, we also sometimes simply maintain the views along with the perceptions that they are justified even when we cannot cite any relevant justification.[254] This is the kind of commitment we would expect if a person were to have an obligatum with a content that ruled out a given action. At the same time, we also clearly engage in the processes that Gibbard calls "normative discussion," and Haidt's research suggests that the dynamics of normative discussion look very much like what Gibbard has proposed. In normative discussion, we often express moral judgments, we give reasons, and we sometimes disagree. We also perceive differences in any attitudes we are expressing as reflecting genuine inconsistencies, which rationally require one or the other of us to revise our opinions. Moreover—and this is important—the psychological research suggests that even if we cannot give each other reasons that the other finds acceptable in these circumstances, we sometimes leave these episodes with more agreement than we began, at least in circumstances where we are discussing things with friends, allies or others with whom we expect to continue interacting. In Haidt's words, the processes of expressing our moral intuitions and trying to give reasons for them in social contexts "exert[s] a constant pressure toward agreement if the two parties were friends and a constant pressure against agreement if the two parties disliked each other."[255] Haidt's research thus suggests that in these contexts, the bare expression of

---

[250] *Id.* at 71-74.

[251] *Id.* at 64-68, 71-80.

[252] *Id.* at 73-75.

[253] *See* Haidt, *supra* note 240.

[254] *Id.* at 819-25.

[255] *Id.* at 820.

conflicting moral intuitions can lead toward consensus. So described, normative discussion is thus a social process, which—if it were the main locus for shifting our moral views in the small hunter-gatherer groups that characterized our environment of evolutionary adaptation—would allow for such shifts to occur while maintaining an appropriate modicum of intrapersonal commitment to shared norms and interpersonal coordination over normative content.

Although the preceding discussion has been limited to moral norms, it should be clear that our legal norms have many of these same features: the content of the law is essentially contestable,[256] and this fact apparently gives us some ability to adapt our legal norms to unanticipated circumstances. If the natural function of the law were to allow us to resolve social contract problems in a flexible manner, we would need psychosocial mechanisms that allow us to adapt our legal norms to changing circumstances as a group while maintaining an appropriate modicum of (i) intrapersonal commitment to the law and (ii) interpersonal coordination over its content. Legal discussion—here defined as the practice of giving and taking reasons concerning what the law is primarily by lawyers and judges in particular instances of adjudication—meets these two criteria. In contexts of adjudication, we express views about what the law requires, we give reasons, and we often disagree. Adjudication nevertheless eventually issues in a final binding judgment that decides the case at hand, even if that judgment cannot be derived from reasons that any particular official antecedently accepted. Importantly, however, the fact that a case has been decided in a final manner does not end the more general question as to what the law is, and it can thus make sense to ask in later cases whether a prior case was decided rightly or what the right bearing of a prior case is on the law.

There are, of course, also many important differences between morality and law. One important one for present purposes is that the law typically contains what Hart called "rules of change," or rules that empower certain officials to legislate and thereby change the law in a more intentional manner than is typical in morality.[257] Rules allowing for legislation are plainly another critical source of adapatibility in our legal codes, and these rules plainly allow us to change the law as a group while maintaining an appropriate modicum of (i) intrapersonal commitment to the law and intrapersonal coordination over its changing content. A fuller account of the relationship between morality and law—including how the two are distinct—would thus need to elaborate just how and why rules allowing for legislation interact with the particular standards that judges employ to identify the law in various societies in ways that can function well, at least under the right social and political conditions. A full account would also need to explain why we would need two distinct coordinating mechanisms and two distinct classes of obligations, and why the law's coordinating mechanism would differ in some ways from morality's. In my view, developing a plausible account of the relevant kind will require a specification of the particular classes of social contract problems that morality and law naturally function to help us resolve, and an elaboration of how these different classes of problems would require coordination mechanisms with these different

---

[256] *See supra* note 2.

[257] *See* HART, *supra* note 5, at 92-93, 95-96,

shapes. I will present my own views on these matters in a subsequent Article, but it would go beyond the scope of this one to try to elaborate that fuller account here. The purpose of this Article is more modest: it is to trace out the deep structure that morality and law share, thus leaving for open the question as to how best to distinguish these normative phenomena.

I. CONCLUSION

If the foregoing arguments are correct, then much of the literature on the law, including many of the descriptive and normative accounts that are familiar from the law and economics literature, have been presupposing a picture of human psychology that is deeply at odds with how we naturally think about obligation. There is now a familiar body of evidence suggesting that we deviate from so called *Homo Economicus*—who reasons only instrumentally—in numerous and systematic ways.[258] The challenge posed here, however, runs deeper. It suggests that the basic social psychological building blocks out of which we create and sustain our moral and legal relations have, and must have, a deep structure that is at odds with current economic frameworks.[259]

Morality and law do not arise from, and could not be sustained only by, separable beliefs about the world and preferences for various states of affairs. Morality and law are instead animated primarily by obligata, which are distinctive portfolios of psychological phenomena that come together—as obligato accompaniments must in any good musical performance—to give morality and law their distinctive lives. Obligata also have a deep structure, the precise contours of which we must learn to understand better. They are what might seem to be a miraculous—and, to my mind, incredibly beautiful—blend of (i) fundamentally agent-centered attitudes toward persons and their motives and actions, along with (ii) attitudes toward shared standards as giving rise to (iii) reasons for action that can (iv) override or exclude many other perceived ones arising from desire or personal interest. Obligata also incline us to (v) react to certain deviations in punitive or critical manners and to deem such reactions warranted or permitted given what the deviations say about how others care about us. But obligata also sensitize us to (vi) the standard excuses, thus allowing us to forgive one another and restore our friendships and relationships despite seeming breaches—at least if the care is real and the seeming breach reveals no genuine lack of concern.

We express obligata in (vii) the special normative terminology that morality and law share, including sometimes in (viii) contexts of discussion or dispute that can become incredibly charged. These interactions naturally engage our attention, and matter to us deeply and inescapably. In these interactions, obligata allow us to (ix) meaningfully disagree, and sometimes reach consensus, even when our resolutions cannot be traced to any particular reasons we antecedently accepted. Obligata are

---

[258] *See, e.g.*, Elizabeth Anderson, *Beyond* Homo Economicus, 29 PHIL. & PUB. AFF 170-200 (2000) (discussing empirical evidence and modern attempts to account for some of it from within basic economic frameworks).

[259] For another set of reasons to question whether our commitments to social norms can be understood instrumentally, see *id*.

nevertheless (x) judgment-sensitive attitudes—in the sense that reasons can be sensibly asked and offered for the judgments we make when we express them—and it is often by this route that we come to terms with one another. Our ability to use this language thus engages (xi) underlying psychosocial mechanisms that can—in the appropriate social and political circumstances—help us maintain sufficient agreement over our sense of what we owe to one another to live well and peaceably together. Obligata thereby give us the capacity to *enjoy* our lives together. Finally, it is possible—though the issue has only been touched upon here and would need to be elaborated further—that our moral and legal judgments (xii) supervene on natural facts because there are natural facts—about what moral and legal rules would conduce to all our objective individual interests in the right way—that partly explain the shape that morality and law take in our lives.

If the arguments in this Article are correct, then the structure of obligata *is* the deep structure of morality and law. Accepting this conclusion would entail seeing that much of the legal literature—including familiar descriptive and normative accounts from law and economics scholars—have been presupposing a psychological picture that is deeply at odds with how we naturally think about obligation. Our capacities to reason instrumentally may not, in fact, figure very centrally at all in our moral or legal practices, and we may necessarily misunderstand these phenomena if we try to shoehorn them into that foreign model. To understand morality and law, we must instead learn to understand better how our distinctive capacities to identify and respond to obligations function.