Integrated Moral Agency and the Practical Phenomenon of Moral Diversity

MICHAEL MOEHLER
Virginia Tech

Abstract: The practical phenomenon of moral diversity is a central feature of many contemporary societies and poses a distinct problem to moral theory building. Because of its goal to settle the moral question fully and exclusively and/or to provide better understanding of moral disagreement, traditional first-order moral theory often does not provide sufficient guidance to address this phenomenon and moral agency in deeply morally diverse societies. In this article, I move beyond traditional first-order moral theorizing and, based on multilevel social contract theory (Moehler 2018, 2020a), develop a practically sound notion of moral agency for morally diverse societies. The interrelational and dynamic notion of *integrated moral agency* developed in this article demands that agents actively exercise their rational and affective capacities, are receptive to the capacities of others, and are aware of the type of moral interaction in which they engage with others. The notion of integrated moral agency helps agents to reconcile conflicting first-order moral directives and to maximally protect agents' autonomy in morally diverse societies.

Keywords: traditional first-order moral theory, multilevel social contract theory, autonomy, integrity, respect

JEL Classification: B15, B25, B31, B41

I. INTRODUCTION

Moral diversity is a central feature of many contemporary societies. In such societies, even after careful consideration of their well-considered moral views, agents often hold divergent moral ideals that cause moral disagreement. If such disagreement is stark and this *practical phenomenon of moral*

diversity (as I call it) endures, then moral diversity does not always facilitate progress, but may lead instead to severe conflict and destructive action. In morally diverse societies, and especially in *deeply morally diverse* societies, the ideal of a fully just society, as judged from the perspectives of all members of society, is often unattainable.

According to Rawls, one central *practical* task of moral and political philosophy is to determine reasoned ways to resolve or, if resolution is not possible, reduce moral and political conflict to ensure mutually respectful, peaceful cooperation:

> One task of political philosophy—its practical role, let's say—is to focus on deeply disputed questions and to see whether, despite appearances, some underlying basis of philosophical and moral agreement can be uncovered. Or if such a basis of agreement cannot be found, perhaps the divergence of philosophical and moral opinion at the root of divisive political differences can at least be narrowed so that social cooperation on a footing of mutual respect among citizens can still be maintained. (Rawls 2001, 2)

According to Rawls, moral diversity matters not only practically, but is also central to moral and political philosophy.[1] One, albeit not very meaningful, way to address the practical phenomenon of moral diversity for moral theory building is to discount its legitimacy. For example, some (although surely not all) moral realists who believe in moral truth and insist that they know it may discount the moral views of others if these views conflict with theirs. Moreover, some moral skeptics who do not believe in morality as traditionally conceived may discount all moral views, including the realists' truth, because moral skeptics do not believe in any legitimate morality. From these viewpoints, the practical phenomenon of moral diversity is either spurious or does not merit further investigation.

At the level of moral theory building, different first-order moral theories often justify conflicting moral conclusions and, in this sense, compete with each other. As Gaus notes:

---

[1] See Gaus (2011, 2016), Muldoon (2016), and Müller (2019).

> We often understand our 'theories of morality' as competing theories describing and explaining the same phenomenon. Indeed, moral philosophers often identify themselves in terms of the adherence to one or the other theory explaining what morality is all about. (Gaus 2011, 551)

First-order moral theory often assumes moral monism, although, as Gaus clarifies, "there is precious little defense of it" (2011, 554). Moreover, if first-order moral theories do allow for plurality of moral conclusions (such as relativistic moral theories), then such theories are typically inadequate for principled resolution of deep moral conflict, precisely because they are too tolerant of diversity at the level of moral theory building.

As a consequence, traditional first-order moral theory typically does not offer sufficient guidance to address the practical phenomenon of moral diversity, although it may help to provide better understanding of the reasons for moral disagreement. Traditional first-order moral theory, in its quest to settle the moral question fully and exclusively, either does not allow sufficient diversity to capture the well-considered moral views of agents, or, if different first-order moral theories together do capture the well-considered moral views of agents, then such theories typically do not offer principled guidance to resolve disagreement among conflicting first-order moral theories. Either way, traditional first-order moral theory typically fails to address the practical phenomenon of moral diversity adequately.

In this article, I do not aim to criticize or discredit the role of traditional first-order moral theorizing, which Scanlon describes as follows:

> All that I have said may seem simply to confirm that, as MacIntyre has written, 'Modern academic philosophy turns out by and large to provide means for more accurate and informed definition of disagreement rather than for progress toward its resolution.' If the 'resolution' in question is a matter of finding arguments that could be deployed to compel agreement between the contending parties, then I agree that philosophy has not been able to provide it and is not very likely to do so. On the other hand, 'more accurate and informed' understanding of disagreement, and of agreement where it exists, seems to me to be an important form of progress—a form that moral theory can reasonably aim at. (Scanlon 1995, 356)

In this article, I move beyond traditional first-order moral theorizing. For a moment, I put on hold—or, depending on one's perspective, build upon the

lessons learned from—the more than 2,500-year-old debate in moral philosophy that aims to determine the ultimately correct view of morality and/or to gain better understanding of moral disagreement. Whatever the conclusion of this debate, if it has one, it will not be able to address the practical phenomenon of moral diversity if agents' well-considered moral views in morally diverse societies, especially in deeply morally diverse societies, are not fully captured by one particular first-order moral theory (which the very existence of different first-order moral theories suggests), and if the precise relationship and jurisdiction of different first-order moral theories are unclear.

I consider the practical phenomenon of moral diversity to be a legitimate concern for moral philosophy. Following Rawls, the core objective of such practical moral philosophy is to determine a moral framework that, despite the ongoing and often severe moral disagreement that is reflected by the conflicting moral conclusions of different first-order moral theories, specifies the moral demands that allow agents to live peacefully with one another on the basis of mutual respect. In the recent literature, several practical (or functionalist) moral and political theories have been developed, in particular by D'Agostino (2003), Gaus (2011, 2016), Muldoon (2016), Müller (2019), Van Schoelandt (2020), and Caton (2020); although not all of these theories explicitly address the practical phenomenon of (deep) moral diversity and develop a notion of moral agency.

Without disregarding other approaches, I build upon multilevel social contract theory (Moehler 2018, 2020a), which offers one possible framework to address the practical phenomenon of deep moral diversity and moral agency. From the perspective of moral theory building, multilevel social contract theory integrates different first-order moral theories. Specifically, in its simplified version, the theory integrates Hobbesian contractarianism, Humean conventionalism,[2] and Kantian contractualism into one systematic moral theory.[3] According to multilevel social contract theory, morality does not consist of one single system of moral rules. Instead, it consists of a multitude of such systems that are valid simultaneously and ordered hierarchically to define the demands of a complex moral world.

---

[2] For discussion of Humean moral conventionalism, see Moehler (forthcoming).

[3] For the difference between 'contractarianism' and 'contractualism', see Darwall (2003).

In this article, I contextualize and systematically develop the *basic features* of the notion of moral agency that underlies multilevel social contract theory.[4] I argue that multilevel social contract theory offers a sound notion of moral agency for morally diverse societies where, especially under the assumption of moral under-determination, moral agency demands the active exercise of agents' rational and affective capacities. It demands that agents are receptive to others and the specific type of moral interaction in which the agents engage with others in order to elicit the most substantial common moral ground. This novel interrelational and dynamic notion of *integrated moral agency* helps to reconcile conflicting first-order moral directives and to maximally protect the autonomy of agents in morally diverse societies. The new notion of integrated moral agency allows agents to make sense of their complex realities in morally diverse societies and offers a principled way to resolve or, if resolution is not possible, mediate moral conflict while treating others with maximal respect.

II. THE STRUCTURE OF MULTILEVEL MORALITY

In its simplified form, multilevel social contract theory (Moehler 2018, 2020a) integrates within one systematic moral theory three different first-order moral theories, namely, 'moral contractarianism' that originates with Hobbes' ([1651] 1996) moral theory, 'moral conventionalism' that originates with Hume's ([1739/1740] 2000) moral theory, and 'moral contractualism' that originates with Kant's ([1785] 1998) moral theory and that has been developed further by Scanlon (1998) and Southwood (2010).[5]

Moral conventionalism and moral contractualism are 'traditional moral theories' (as I employ the term). Traditional moral theories assume, as a basis for the justification of moral rules, that agents value moral ideals (shared or not) at least partially for intrinsic reasons or embrace such ideals for other traditional moral reasons, such as altruistic reasons or similarly motivated other-regarding reasons. As a result, traditional moral theories cannot fully capture the practical phenomenon of deep moral diversity that includes agents who do not embrace morality on traditional, noninstrumental

---

[4] Related but conceptually different notions of moral agency have been developed by defenders of Kantian constructivist ethics, in particular Korsgaard (2009).
[5] See Gauthier (1997, 134–135), Watson (1998, 173–174), Darwall (2003, 1–8), D'Agostino, Gaus, and Thrasher (2017), and Moehler (2018, 11–12; 2020a, 3–7).

grounds. Moral contractarianism, by contrast—as a result of its purely instrumental approach—can accommodate deep moral diversity, and thus can complement moral conventionalism and moral contractualism from the perspective of moral theory building so long as agents share the overarching goal of ensuring peaceful cooperation.

Multilevel social contract theory represents a distinct position in moral theory that, even in its simplified version, differs from Parfit's (2011) triple theory. Parfit's theory holds that rule consequentialism, Kantian contractualism, and Scanlonian contractualism lead to similar moral conclusions and thus represent different ways to 'climb the same mountain'. Multilevel social contract theory, by contrast, combines three different contractarian moral theories within one systematic moral theory that entails Humean, Hobbesian, and Kantian moral features. Multilevel social contract theory does not claim that different contractarian moral theories converge to reach similar moral conclusions, although it suggests so in a weak sense with respect to Hobbes' and Kant's moral theories (as I clarify in section III). Instead, multilevel social contract theory considers different contractarian moral theories to be valid for different domains of morality.

Multilevel social contract theory purports that, in order to climb a mountain successfully, different types of moral theory may apply the higher one climbs. If, with increasing height, the air becomes thinner and the moral terrain more demanding, then different moral theories may be needed that are valid for these different circumstances and may prescribe different behaviors (what one ought to do). The higher one climbs and the thinner the air becomes, then, morally speaking, the less specific the demands of morality may be, although the demands need not necessarily be less significant. If one comes close to the top of the mountain where, after reaching the point of no return, the air is so thin that, morally speaking, no moral basis as traditionally conceived exists any longer among agents apart from the agents' goal to reach the top of the mountain, then pure instrumental morality is the only guide left to ensure one's survival.

Multilevel social contract theory also differs from Southwood's (2010) moral theory, which holds that existing contractarian moral theories, especially Hobbesian contractarianism and Kantian contractualism, do not offer plausible moral foundations. According to Southwood, Hobbesian contractarianism:

> At best appears to get morality wrong in the right way […] on account of its reliance on an implausibly personal and partial characterization of the moral point of view. Kantian contractualism at best appears to get morality right in the wrong way […] on account of its reliance on a substantive conception of practical reason. (Southwood 2010, 190)

Southwood's (2010, 88–96, 124–128) 'deliberative contractualism' represents an intermediate position that, according to this view, is superior to Hobbesian contractarianism and Kantian contractualism, and demands that agents actively engage in deliberation with others to reach consensus on a common code by which to live.

Multilevel social contract theory, by contrast, keeps the core features of Hobbesian moral contractarianism, Humean moral conventionalism, and Kantian moral contractualism intact, and, in doing so, maintains the strengths of each of these theories. In order to capture the practical phenomenon of moral diversity and offer a principled way to resolve conflict in deeply morally diverse societies, multilevel social contract theory considers the moral domain to be heterogeneous. It assumes that agents may agree on different social contracts that are valid simultaneously for different types of moral interaction and that such contracts may be restricted in their universality and generality. Nevertheless, it requires that agents are fit for social cooperation and morality, and will ultimately agree on regulations for all relevant types of moral interaction to ensure mutually respectful, peaceful cooperation.

Moreover, multilevel social contract theory assumes that moral rules are ordered hierarchically, with lower-level moral rules taking priority over higher-level moral rules in the regulation of moral interactions. The theory assumes that agents will always justify moral rules to themselves and others based upon the most substantial common moral ground with others, both in the traditional and purely instrumental understandings of morality, for the most local domain. The theory assumes that, based on some common denominator as a starting point or an end point for the justification of moral rules that may vary for different types of moral interaction, agents will always agree on the least invasive system of moral rules to regulate their moral interactions. In the following, I clarify the specific nature of multilevel morality that determines the basic features of its underlying notion of moral agency.

III. THE NATURE OF MULTILEVEL MORALITY

In keeping with the plurality thesis, the practical view of morality defended by multilevel social contract theory does not aim to capture all aspects of morality, but focuses on *social* morality as opposed to *personal* morality. Specifically, the goal of multilevel morality is to harmonize moral interactions among agents while ensuring peaceful cooperation independent of the precise origin of morality, such as evolution from actual social practices or justification by rational procedures or both.[6] Nevertheless, one core challenge for multilevel social contract theory (to the extent that the practical view of morality underlies this challenge) is to show that multilevel morality, especially for the domain of pure instrumental morality, defends a genuine form of morality because, under certain conditions, the demands of pure instrumental morality are assumed to override the demands of traditional morality.[7]

The first aspect of this challenge concerns moral motivation and is expressed by Prichard's dilemma.[8] Prichard argues that there is no good reason to act morally. If one refers to moral reasons, such as altruistic reasons, then one presupposes the persuasive force of morality, which is circular reasoning. If one refers to nonmoral reasons, such as self-interest, then one provides the wrong kind of reasons to act morally. In contractarian moral theory, Scanlon's contractualism provides a direct response to Prichard's dilemma. Scanlon (1998, 147–158) argues that there is a third type of reason to act morally that is neither a moral reason nor a nonmoral reason as traditionally conceived, but that stems from the consideration of an action's wrongness and its justifiability towards others. According to Scanlon, this account of moral motivation:

> Is closely enough connected to our ideas of right and wrong to be clearly an account of *moral* reasons, but it is not so closely identified with these ideas as to amount to the trivial claim that the reason we have to avoid certain actions is just that they are wrong. (Scanlon 1998, 187)

---

[6] Despite this feature, genuine social moral rules, as understood here, are considered to be distinct from merely 'socially constructed norms' that may or may not have moral significance. For discussion of socially constructed norms and their relationship to moral theory, see Valentini (2021). For discussion of potential limitations of this view of morality, see Morris (2020) and Moehler (2020b, 97–98).

[7] For related discussion, see Moehler (2020a, 15–16).

[8] See Prichard (1912).

Even if correct, Scanlon's contractualist moral theory is valid only for agents who *de facto* are "moved by the aim of finding principles that others, similarly motivated, could not reasonably reject" and who are not moved merely by "seeking some kind of advantage" (1998, 5). Although Scanlon (1998, 158–160) highlights the importance of agents caring about such justification, Scanlon's theory, as a result of its contractualist features, cannot sufficiently address the practical phenomenon of moral diversity that includes agents who, at least in certain types of moral interaction, may exclusively seek their own advantage despite sharing a common goal with others. Such agents fall outside the scope of Scanlon's moral theory. This finding does not constitute a criticism of Scanlon's theory which, as a result of its traditional moral nature, pursues a different task.

Multilevel social contract theory, within the domain of pure instrumental morality, includes such agents who may exclusively seek their own advantage. Yet, the theory does not fall prey to what Southwood (2019) calls the 'concessional fallacy', namely that a moral theory may not be sufficiently demanding because it overly caters to agents' self-interests. Multilevel social contract theory relies on the 'principle of subsidiarity', and thus justifies moral rules always on the most substantial traditional moral ground (shared or not). The theory defers to pure instrumental reasoning only in cases where traditional morality fails. In Hobbesian terms, multilevel social contract theory takes agents 'as they are' as a basis for morality, although the thinner the traditional moral ground among agents, the more carefully agents must reflect on their traditional moral views and their importance compared to the agents' aim of reaching their overarching goal.

Because multilevel social contract theory, for the domain of pure instrumental morality, does not require that agents follow moral rules on the basis of what are traditionally conceived to be moral reasons (although it does not rule out such reasons) but allows agents to be motivated entirely by self-interest, the theory faces the 'wrong kind of reasons' objection. More fundamentally, the problem concerns the priority of prudential moral reasoning over traditional moral reasoning for the domain of pure instrumental morality.[9] To be clear, within the domain of pure instrumental morality, multilevel social contract theory allows agents to value their traditional moral views highly. It generally only rules out that agents value their traditional moral

---

[9] See Gaus (2019, 110–111).

views infinitely over other agents' traditional moral views, their lives, and the general goal of ensuring peaceful cooperation.

Such agents, whom I call *homo categorical*,[10] are not willing to make concessions in conflict situations, which puts them deeply at odds with others and renders mutually respectful peaceful cooperation impossible in societies where the practical phenomenon of moral diversity endures. Such agents, often as a result of ideological considerations, are dogmatic and fall outside the scope of pure instrumental morality. Also, if agents, such as extreme glory seekers or suicide bombers, do not embrace the overarching goal to ensure peaceful cooperation, including an interest in preserving their lives and the means necessary to do so, then they fall outside the scope of the theory.[11] Even if pure instrumental morality imposes only weak normative constraints on the reasoning of agents, the theory has normative force and thus considers some agents to be potentially irrational.

For cases of conflict in which moral reasoning is reduced to instrumental reasoning and peace is at risk, multilevel social contract theory assumes that the demands of pure instrumental morality override the demands of traditional morality. In societies in which the practical phenomenon of moral diversity endures, by definition agents disagree on what is morally right as traditionally conceived and what qualifies as genuine moral reasons. Under the assumption of deep moral diversity, traditional moral reasons are just one type of reason that may or may not help to harmonize the behavior of agents. Under such conditions, traditional moral reasons are not privileged reasons and may often be the main cause for conflict.

In cases where pure instrumental morality applies, multilevel social contract theory assumes that life, and the human cooperation that is necessary to sustain it, are more important than traditional morality. Or, traditional morality is misconceived if it pits agents against each other and endangers their existence.[12] In such cases, agents must distance themselves from their traditional moral views to evaluate their overall interests and mediate conflicts according to the demands of pure instrumental rationality to ensure that, in Rawls' terms, "social cooperation on a footing of mutual respect among citizens can still be maintained" (2001, 2). According to multilevel

---

[10] See Moehler (2018, 103).
[11] See Moehler (2020b).
[12] See Moehler (2019, 145).

morality, agents who are not able to do so are not fit for cooperation and social morality. Multilevel social contract theory does not claim that there is no space outside of morality. Instead, it merely extends the limits of morality as far as methodologically possible within the bounds of social morality and moral agency.

The second aspect of the challenge to show that multilevel morality represents a genuine form of morality concerns the role and nature of moral emotions. Arguably, pure instrumental morality does not sufficiently consider the importance of agents' affective capacities, such as anger, indignation, guilt, blame, and resentment, and more generally the moral psychology of agents. Moreover, as Southwood (2008, 185–186; 2010, 34–42) stresses, even if pure instrumental morality were to consider such capacities, it may misconstrue their nature as primarily self-directed. To be clear, the fact that multilevel social contract theory prioritizes reason over emotion within the domain of pure instrumental morality does not imply that the theory neglects the importance of such traditional moral concepts. Instead, it merely considers them to be typically part of traditional morality. According to multilevel social contract theory, both reason and emotion have their roles in moral agency. Nevertheless, in situations where traditional morality turns agents against each other and threatens their existence, traditional moral concepts cannot serve as the ultimate moral compass because they are often the very cause for conflict.

The third aspect of the challenge concerns the content of morality. In order to be considered legitimate, the moral rules that are justified by pure instrumental morality must, in the relevant sense, resemble traditional moral rules. Although this consideration is not essential for my argument here, the moral rule that I defend for the domain of pure instrumental morality under the specific circumstances described (under different circumstances, different rules may be justifiable) fulfills this condition.[13] I call this rule the *weak principle of universalization* and argue that it can be considered to be a weak version of Kant's categorical imperative, because it weakly expresses the moral ideals that underlie Kant's moral law. Nevertheless, the rule is not as general and universal as Kant's categorical imperative because it applies only to specific types of conflict and only to agents in this empirical world. The advantage of the weak principle of universalization is that

---

[13] See Moehler (2018, 133–139; 2020a, 48–52).

agents do not need to embrace these moral ideals on traditional moral grounds, because these ideals can be justified instrumentally. Ultimately, multilevel social contract theory stresses the respect that agents owe each other if they want to coexist peacefully despite the endurance of the practical phenomenon of moral diversity. Based on its general assumptions, multilevel social contract theory defines the bounds of social morality and thus also the bounds of meaningful moral agency.

IV. AUTONOMY, INTEGRITY, AND AGENCY

Multilevel social contract theory, within its defined bounds, aims to provide agents with maximal autonomy in morally diverse societies. To this end, the theory allows several levels of agreement among agents on moral rules that may be restricted with respect to their universality and generality—within both the domain of traditional morality and that of pure instrumental morality, as well as within the domain of what Van Schoelandt aptly calls "intermediate moralities" that combine features of both types of morality (2019, 133).[14] Although the discussion in this article focuses primarily on triple theory that combines Hobbesian contractarianism, Humean conventionalism, and Kantian contractualism, multilevel social contract theory allows for further levels of moral rules that define a more fine-grained n-level social contract theory and that determine a complex web of moral relationships in diverse societies.

In order to allow for moral diversity, multilevel social contract theory does not require that potential demands stemming from nonmoral interpersonal relationships must be consistent with the demands of morality for the entire moral domain, as, for example, Scanlon's traditional moral theory requires. According to Scanlon's theory, morality defines the most fundamental demands on human agency, and any additional nonmoral social obligations that may arise through love, friendship, and other social relationships, if adequately conceived, must respect the priority of traditional morality. Such demands must be sensitive to the demands of traditional morality. As Scanlon puts it:

---

[14] For further discussion of the features of traditional morality related to compliance issues, see Van Schoelandt (2019, 132).

Interacting with others *qua* chess players, *qua* lovers, or poets, is a special case of interacting with them *qua* rational creatures. If we didn't think of them as having the status of rational creatures, we wouldn't be able to relate to them in the way that we do. Therefore, I would say that a relationship to others that brings the moral requirements of 'what we owe to each other' in train is presupposed by the more specific forms of relationships. That is one of the reasons why moral requirements take precedence over other relationships. (Scanlon and Voorhoeve 2001, 30)

According to Scanlon's theory, a person who murders another person for a friend is not a real friend. Scanlon considers the priority of traditional morality and its presupposition by other interpersonal relationships to be a necessary condition for the creation of a kingdom of ends in the Kantian sense.

Also, in the context of his discussion of moral relativism, Scanlon stresses that, although his theory allows for variable moral practices (whereby an action that is wrong in one context may be morally unobjectionable in another context), ultimately the moral force of such practices "is explained by appeal to a single substantive moral principle", a position that Scanlon calls "parametric universalism" (1998, 340). Following Kant, Scanlon's moral theory assumes that the categorical imperative is valid for the whole moral domain and thus any lower-level moral rules that are expressed in the form of hypothetical imperatives, which do not qualify as genuine moral rules in Kant's moral theory, cannot override or run counter to the demands of the categorical imperative from the perspective of moral agency.[15]

Multilevel social contract theory, by contrast, does not require that the demands of different levels of morality are consistent with each other, because conflicts within the domain of lower-level morality may often be the very reason for appeal to higher-level morality.[16] For justificatory purposes, according to multilevel social contract theory, the normative content of higher-level morality is independent of the normative content of lower-level morality. According to multilevel social contract theory, higher-level morality has moral authority precisely in situations where lower-level morality has

---

[15] See Kohl (2018).
[16] See Moehler (2018, 21).

failed to resolve conflict. In agreement with Scanlon's theory, multilevel social contract theory considers morality to be foundational. However, in contrast to Scanlon's theory, multilevel social contract theory considers morality also to be layered and hierarchical.

As such, within the domain of pure instrumental morality, a plurality of reasons—including reasons that stem from personal commitments, such as friendship, love, and other social relationships—may be part of the justificatory basis for moral rules. Furthermore, the demands of such personal reasons need not be consistent with the demands of traditional morality because, within the domain of pure instrumental morality, traditional moral reasons themselves may be controversial and the primary cause for conflict. Within the domain of pure instrumental morality, traditional moral reasons are not privileged reasons. Instead, agents' nonmoral reasons as traditionally conceived may be just as important as their traditional moral reasons for ensuring mutually respectful, peaceful cooperation.

The diversity of reasons that multilevel social contract theory allows within the domain of pure instrumental morality helps to protect the integrity of agents, which, according to Bernard Williams, is an important feature of a sound moral theory (to the extent that Williams can be considered to be interested in moral theory). In his critique of utilitarianism, Williams (1973, 100–118) argues that agents' lives are structured around people and projects they care about, including their own wellbeing, families, friends, and the fulfillment of basic necessities of life, as well as cultural and aesthetic interests. Such relationships and projects, and the commitments that follow from them, give meaning to agents' lives and constitute their identity as persons. Integrity demands that agents' identities are protected. Williams argues that utilitarianism, because of its focus on the group level, does not sufficiently protect the integrity of agents. Utilitarianism demands strict impartiality and impersonality of agents with regard to their own projects and the projects of others, which may alienate agents from their own lives.[17] A sound moral theory must provide room for personal relationships and the pursuit of agents' important life projects.

Scanlon's moral contractualism responds to this criticism. According to Scanlon (1998, 160), the morality of 'what we owe to each other' provides

---

[17] See also Moehler (2013).

agents with sufficient room to pursue personal relationships and life projects, because moral rules that do not allow such personal space could be reasonably rejected by agents. That is, although Scanlon's moral theory assigns priority to traditional moral reasons as the foundation for social relationships, his theory leaves sufficient room for personal considerations. However, this provision holds only if all members of society are reasonable in the specific way presupposed by Scanlon's theory, because the demands of the concept of reasonableness determine the claims to which agents can object.[18] In morally diverse societies, the demands of the concept of reasonableness are likely to be controversial and, in deeply morally diverse societies, not all agents can be assumed to be reasonable. As a result, Scanlon's moral theory, which is limited to the bounds of traditional moral reasoning, may compromise the integrity of agents who disagree with the prevalent traditional morality. The interests of such agents are excluded from Scanlon's moral theory.

Multilevel social contract theory, by contrast, includes the interests of such agents, in cases where traditional morality fails to resolve moral conflict. Within the domain of pure instrumental morality, not only traditional moral reasons can serve as reasons for the justification of moral rules, but also other reasons that may stem from important personal relationships and life projects. Moral theory, if it aims to address the practical phenomenon of moral diversity, must consider such reasons to which agents may be strongly committed. Even if multilevel social contract theory limits the bounds of social morality and moral agency (as discussed in section III), the theory helps to maximally protect the integrity of agents in deeply morally diverse societies where traditional moral reasons are controversial.

Moreover, multilevel social contract theory suggests a complex notion of moral agency that entails that, in deeply morally diverse societies, agents may need to follow different moral rules for different types of moral interaction. To illustrate this idea in a nonmoral context, Nguyen (2019, 2020) suggests in his analysis of 'games' (in their broadest sense) and the 'rules' that govern them that games do not merely create structures for social interaction, but that they also create 'temporary selves'. Games can be seen as describing (sometimes arbitrary) rules for different types of social interaction that shape agents' experiences. Nguyen's analysis of social interaction

---

[18] For further discussion of this consideration, see Suikkanen (2019).

through the lens of games shows that agency is typically not only socially embedded, modular, and fluid, but also often requires that agents adopt, at least temporarily, certain ends for successful social interaction. Moreover, playing different games with different rules may allow agents to understand more fully their own selves and the importance of constraints for personal and social development.

The notion of moral agency that is implicit in multilevel social contract theory does not hinge on Nguyen's analysis. Nevertheless, for some readers, Nguyen's analysis may be helpful for better understanding the complex notion of moral agency that is suggested by multilevel social contract theory. Multilevel social contract theory—with its different levels of morality that define a complex web of moral rules that are valid simultaneously for different types of moral interaction—suggests a socially embedded, differentiated, and fluid notion of agency that is similar to the one found in Nguyen's analysis. One central difference to Nguyen's analysis is that multilevel social contract theory focuses exclusively on the domain of morality and thus underlies the (non-arbitrary) constraints of moral theory. In the following, I lay out the basic features of the notion of integrated moral agency that underlies multilevel social contract theory.

V. INTEGRATED MORAL AGENCY

As the previous discussion of the structure and nature of multilevel social contract theory implies, the notion of moral agency that is defended by multilevel social contract theory differs from other notions of moral agency by *integrating* different contractarian moral theories within one systematic moral theory.[19] Multilevel social contract theory integrates traditional moral reasoning, as captured by Humean moral conventionalism and Kantian moral contractualism, with pure instrumental or prudential moral reasoning, as captured by Hobbesian moral contractarianism.[20] Because most readers will be familiar with traditional moral reasoning, I will focus primarily on the aspects of moral agency that pertain to the domain of pure instrumental morality and their integration with traditional morality.

---

[19] For discussion of other notions of moral agency, especially notions of moral agency developed by supporters of Kantian constructivist ethics, see Korsgaard (2009).
[20] For Hobbes' prudentialism, see Abizadeh (2018, 110).

For the domain of pure instrumental morality, multilevel social contract theory relies on a particular Hobbesian model of moral agency that I call the 'homo prudens model'.[21] The homo prudens model assumes that, in moral interactions in which traditional moral reasoning fails and pure instrumental rationality is the only means to ensure peace, agents, in addition to reflecting upon their goals and empirical conditions, are forward-looking and value their lives and the expected gains from peaceful cooperation more than they value noncooperation per se.

The homo prudens model is a close cousin of the homo economicus model (the predominant model of human agency in economic theory) that is often thought to represent a particularly liberal notion of human agency. According to Gauthier (1986, 345), instrumental rationality, especially in the way it is employed by Gauthier in the context of his theory of 'morals by agreement', captures the reasoning and affections of economic agents and, more specifically, the affections of the 'liberal individual'. Gauthier (1986, 353–354) argues that the liberal individual is only a recent invention and that it is unclear whether or not her 'ecological niche' can be realized and sustained. As such, Gauthier's moral theory may be bound by time.

This characterization of the homo economicus model does not apply to the homo prudens model employed by multilevel social contract theory. Multilevel social contract theory assumes that instrumental rationality is an essential and permanent part of human agency because the capacity for instrumental reasoning is generally necessary for agents to survive in this empirical world. In this empirical world, instrumental reasoning can be viewed as a timeless feature of human agency. Moreover, according to multilevel social contract theory, the homo prudens model represents a partial model of human agency that is assumed to be valid only for moral interactions in which moral reasoning is reduced to instrumental reasoning and instrumental morality is the only means to ensure peace.

In other words, agents' behavior within the domain of traditional morality is not assumed to be guided by the demands of the homo prudens model and, more specifically, the ideal of individual utility-maximizing behavior. Instead, as discussed in section III, such behavior is assumed to be guided by traditional morality and its core notions, including moral emotions and agents' affective capacities that, in the specific case of moral contractualism,

---

[21] See Moehler (2018, 95; 2020a, 48).

may have a certain liberal content. The *homo prudens* model does not aim to capture the whole range of moral behavior, but only the behavior relevant to certain types of moral interaction. As such, depending on the specific type of moral interaction, one and the same agent may be guided sometimes by traditional morality and sometimes by the *homo prudens* model. According to multilevel social contract theory, different types of moral interaction may require different constructions of morality, depending on the common moral ground among agents, both in the traditional and purely instrumental understandings of morality.

This feature of multilevel social contract theory does not imply that agents are assumed to have a split personality. Instead, multilevel social contract theory merely assumes that, under the endurance of the practical phenomenon of moral diversity, agents consider all relevant reasons for the justification of moral rules in the light of their overarching goal of ensuring peaceful cooperation. Because peaceful cooperation is an interdependent variable and because, in the domain of pure instrumental morality, agents are a potential threat to each other, agents may need to abstract from certain aspects of traditional morality—at least in certain types of moral interaction—in order to reach their overarching goal of ensuring peaceful cooperation (as conceptually expressed by the restriction of the generality of moral rules). For the justification of moral rules, agents must consider their goals and moral views as well as those of other agents, even if agents are not intrinsically motivated to justify their actions towards others, as assumed by traditional moral theories.

Further, according to multilevel social contract theory, moral agency is *interrelational*. According to multilevel social contract theory, moral demands depend not only on the specific type of moral interaction, but also on the agents' cooperative partners (as conceptually expressed by the restriction of the universality of moral rules). Because multilevel social contract theory demands that agents always determine the most substantial common moral ground with their cooperative partners, in terms of both the traditional and purely instrumental understandings of morality, sometimes agents may appeal to their full-fledged traditional morality, or certain parts of it, and sometimes they may need to rely on pure instrumental morality.

Stated differently, agents may be bound by their traditional morality with regard to some agents and/or some types of moral interaction and by pure

instrumental morality with regard to other agents and/or other types of moral interaction. Agents may be under the authority of different moralities (or intermediate moralities) at the same time, depending on their cooperative partners and the precise form(s) of moral interaction. Agents may be in multiple moral relationships with each other that require different forms of moral reasoning. In this sense, agents may be considered to engage in moral role-playing, which demands that agents take on different moral perspectives and reconcile conflicting first-order moral directives by accepting different moral constraints (rules) for different agents and/or different types of moral interaction. Agents may be considered to play different 'moral games' that are governed by different moral rules.

Phenomenologically, engaging in such moral role-playing may provide agents with a better understanding of their own selves and other agents, which may lead agents to a less rigid and more constructivist understanding of morality. In addition, perceiving moral agency in such an integrated and interrelational manner allows for differentiation of agents' moral views, which facilitates the reconciliation of conflicting moral views both within and among agents, especially in morally fragmented societies that are prone to conflict.[22] Moral differentiation, which is inherent in the structure of multilevel social contract theory, requires that agents consider the universality and generality of their own moral views and the moral views of others, as well as reflect upon the origin and legitimacy of such views and their underlying values.

Moral differentiation may render it more acceptable for agents to make concessions with regard to their moral views as temporary means for social interaction, because at no point is an agent's entire worldview at stake. Moral differentiation may also help to preserve agents' integrity in morally fragmented societies, and thus to address Williams' (1973) integrity concern discussed in the previous section, even if, as Williams (1985) concedes, not all human projects may ultimately be fully reconcilable.[23] The developed notion of moral agency allows agents to connect their moral views systematically to

---

[22] For discussion of this point, see Ostrom and Ostrom (2002, 96). For discussion of the notion of fragmentation in the context of collective moral responsibility, see Braham and van Hees (2018).

[23] For critical discussion of Williams' view in the context of virtue ethics, see Cottingham (2010).

form a unified, integrated moral self, even if the notion of moral agency demands that, due to the complexities of deeply morally diverse societies, agents carefully differentiate their moral views.

Moreover, the integrated and interrelational notion of moral agency that is defended by multilevel social contract theory encourages agents to perceive disagreements as opportunities for discovery, learning, and change. It invites agents to probe their views both conceptually and historically and, if they consider the moral views of others to be legitimate, to revise their views and construct the best moral world together based on the agents' combined perspectives. The notion of moral agency asks agents to exercise their moral capacities as an integral part of the determination of the constraints of morality that agents impose on themselves and others. Such active engagement with others may allow agents to become better moral agents over time because it requires that the agents carefully reflect on their views and the moral views of others, even if disagreements remain and consensus is a false ideal in morally diverse societies.

Returning to the distinction between traditional and purely instrumental morality, in practice, there will probably be a continuum with regard to agents' moral reasons and motivations that involves, at the most foundational level, traditional moral concepts and that ascends to pure instrumental morality. As a result, in practice, agents' moral reasons and motivations may not be captured fully by moral conventionalism, moral contractualism, and moral contractarianism. As indicated in section II, the division of contractarian moral theory into three types of moral theory is intended merely to clarify the core differences among different contractarian moral theories. In practice, under the phenomenon of moral diversity, moral agency is typically more complex. Multilevel social contract theory can account for such complexity that determines a complicated web of moral relationships among agents within the domains of both traditional and pure instrumental morality.

A final feature of multilevel social contract theory stems from the fact that it regards traditional morality to be culturally-dependent and path-dependent, because the theory regards traditional morality to be grounded at least partially in evolving social moral practices, whereas the demands of pure instrumental morality are assumed to be based upon general and fairly stable assumptions about human nature, empirical facts, and conditions of

rationality. Nevertheless, in the form discussed, the demands of multilevel morality are valid only for moral interactions among agents who live in this empirical world and who reason in the mode of *homo prudens*. This feature of pure instrumental rationality stands in contrast to Kant's moral theory, which assumes that moral rules carry with them absolute necessity and thus are valid for all rational beings in the universe.[24] As a result of this feature, the precise moral demands of multilevel social contract theory cannot be determined *a priori*, but are constructed by the exercise of practical reasoning in moral interactions.

In this sense, multilevel social contract theory suggests a *dynamic* notion of moral agency. As a result of combining evolutionary aspects of morality in the domain of traditional morality with static (although not necessarily eternally fixed) rationalistic aspects in the domain of pure instrumental morality, the notion of moral agency defended by multilevel social contract theory requires that agents (re)construct the demands of morality under evolving social conditions by exercising their rational and affective capacities. Multilevel social contract theory takes seriously the primacy of evolved morality and its relationship to reason and thus the lessons of Gaus' (2018a, 2018b, 2018c, and forthcoming) work on social morality. Nevertheless, one essential difference between multilevel social contract theory and Gaus' moral theory is that, in the face of moral diversity, multilevel social contract theory, which follows Hume's notion of evolutionary processes, combines evolution and reason *hierarchically*, whereas Gaus' theory, which follows Hayek's (1960, 1973) notion of evolutionary processes, combines them *sequentially*.

Also, whereas both theories respond to the challenge of moral diversity by defending decentralized, dynamic processes of moral explanation and justification, Gaus' (2011, 2016) moral theory makes room for moral diversity by dispensing with the ideal of justifying specific moral rules. Multilevel social contract theory, by contrast, makes room for moral diversity by integrating different contractarian moral theories and consideration of their adequate scope. Conceptually, multilevel social contract theory suspends the requirement of uniqueness for lower-level morality to allow for moral diversity. The theory defends uniqueness only when it is necessary to ensure principled conflict resolution after all other moral means have failed.

---

[24] See Kant ([1785] 1998, AK 4:389).

Overall, the notion of moral agency that multilevel social contract theory defends is complex, and demands the active exercise of agents' rational and affective capacities. Multilevel morality demands that agents are receptive to others and the specific type of moral interaction in which they engage so that agents always elicit the most substantial common moral ground for their moral interactions, especially under conditions of moral uncertainty and under-determination.[25] In practice, in order to avoid that agents are unable to act because too many considerations apply, which Scanlon (1998, 170) calls 'moral gridlock', or because too much uncertainty prevails, agents may idealize some of the features of their empirical environment and/or their counterparts, especially if further information is unavailable and inaction would cause significant harm. Also, in practice, moral heuristics may evolve that can serve as shortcuts for behavior in a morally fragmented world.

To conclude, because its goal is to settle the moral question fully and exclusively and/or to provide better understanding of the reasons for moral disagreement, traditional first-order moral theory often does not offer sufficient guidance to address the practical phenomenon of moral diversity and moral agency in contemporary societies. Multilevel social contract theory can address this phenomenon and, in doing so, it reorients moral theory. It reorients moral theory to be practical and accept that, in a morally diverse society, no single correct system of moral rules exists. Instead, many such systems work together simultaneously and hierarchically to protect the autonomy of agents maximally in morally diverse societies. As a result of its structure and nature, multilevel social contract theory suggests a practically sound notion of moral agency for morally diverse societies. In this article, I have systematically placed into context and developed the basic features of this notion of moral agency, even if space restrictions do not allow discussion of all its aspects. The developed interrelational and dynamic notion of integrated moral agency suggests one particular way for agents to make sense of their complex realities in morally diverse societies and offers guidance with regard to treating others respectfully in a complex moral world.

---

[25] For discussion of some aspects related to ethical decision-making under risk and uncertainty, see Rowe and Voorhoeve (2018) and Rowe (2019).

REFERENCES

Abizadeh, Arash. 2018. *Hobbes and the Two Faces of Ethics.* Cambridge: Cambridge University Press.

Braham, Matthew, and Martin van Hees. 2018. "Voids or Fragmentation: Moral Responsibility for Collective Outcomes." *The Economic Journal* 128 (612): F95–F113.

Caton, James. 2020. "Moral Community and Moral Order: Developing Buchanan's Multilevel Social Contract Theory." *Erasmus Journal for Philosophy and Economics* 13 (2): 1–29.

Cottingham, John. 2010. "Integrity and Fragmentation." *Journal of Applied Philosophy* 27 (1): 2–14.

D'Agostino, Fred. 2003. *Incommensurability and Commensuration: The Common Denominator.* Aldershot: Ashgate.

D'Agostino, Fred, Gerald Gaus, and John Thrasher. 2017. "Contemporary Approaches to the Social Contract." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Article published March 3, 1996; last modified May 20, 2021. https://plato.stanford.edu/entries/contractarianism-contemporary/.

Darwall, Stephen L., ed. 2003. *Contractarianism/Contractualism.* Oxford: Blackwell Publishing.

Gaus, Gerald. 2011. *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World.* Cambridge: Cambridge University Press.

Gaus, Gerald. 2016. *The Tyranny of the Ideal: Justice in a Diverse Society.* Princeton, NJ: Princeton University Press.

Gaus, Gerald. 2018a. "It Can't Be Rational Choice All the Way Down: Comprehensive Hobbesianism and the Origins of the Moral Order." In *Buchanan's Tensions: Reexamining the Political Economy and Philosophy of James M. Buchanan*, edited by Peter J. Boettke and Solomon Stein, 117–147. Arlington, VA: Mercatus Center.

Gaus, Gerald. 2018b. "The Priority of Social Morality." In *Morality, Governance, and Social Institutions: Reflections on Russell Hardin*, edited by Thomas Christiano, Ingrid Creppell, and Jack Knight, 23–56. Cham: Palgrave McMillan.

Gaus, Gerald. 2018c. "Self-Organizing Moral Systems: Beyond Social Contract Theory." *Politics, Philosophy & Economics* 17 (2): 119–147.

Gaus, Gerald. 2019. "Moral Conflict and Prudential Agreement: Michael Moehler's *Minimal Morality*." *Analysis* 79 (1): 106–115.

Gaus, Gerald. Forthcoming. *The Open Society and Its Complexities.* New York, NY: Oxford University Press.

Gauthier, David. 1986. *Morals by Agreement.* Oxford: Clarendon Press.

Gauthier, David. 1997. "Political Contractarianism." *The Journal of Political Philosophy* 5 (2): 132–148.

von Hayek, Friedrich A. 1960. *The Constitution of Liberty.* Chicago, IL: University of Chicago Press.

von Hayek, Friedrich A. 1973. *Law, Legislation and Liberty. Volume 1: Rules and Order.* Chicago, IL: University of Chicago Press.

Hobbes, Thomas. (1651) 1996. *Leviathan.* Edited by Richard Tuck. Cambridge: Cambridge University Press.

Hume, David. (1739/1740) 2000. *A Treatise of Human Nature.* Edited by David Fate Norton and Mary J. Norton. Oxford: Oxford University Press.

Kant, Immanuel. (1785) 1998. *Groundwork of the Metaphysics of Morals.* Translated and edited by Mary Gregor. Cambridge: Cambridge University Press.

Kohl, Markus. 2018. "Kant's Critique of Instrumental Reason." *Pacific Philosophical Quarterly* 99 (3): 489–516.

Korsgaard, Christine M. 2009. *Self-Constitution: Agency, Identity, and Integrity.* Oxford: Oxford University Press.

Moehler, Michael. 2013. "Contractarian Ethics and Harsanyi's Two Justifications of Utilitarianism." *Politics, Philosophy & Economics* 12 (1): 24–47.

Moehler, Michael. 2018. *Minimal Morality: A Multilevel Social Contract Theory.* Oxford: Oxford University Press.

Moehler, Michael. 2019. "Replies to Gaus, Van Schoelandt and Cooper: Prudence, Morality and the Social Contract." *Analysis* 79 (1): 140–153.

Moehler, Michael. 2020a. *Contractarianism.* Cambridge: Cambridge University Press.

Moehler, Michael. 2020b. "Minimal Morality, Bargaining Power, and Moral Constraint: Replies to D'Agostino, Thrasher, Morris, and Vanderschraaf." *Analytic Philosophy* 61 (1): 87–100.

Moehler, Michael. Forthcoming. "Strategic Justice, Conventionalism, and Bargaining Theory." *Synthese.*

Morris, Christopher W. 2020. "Morality's Many Parts: Symposium on Michael Moehler's *Minimal Morality.*" *Analytic Philosophy* 61 (1): 57–69.

Muldoon, Ryan. 2016. *Social Contract Theory for a Diverse World: Beyond Tolerance.* New York, NY: Routledge.

Müller, Julian F. 2019. *Political Pluralism, Disagreement and Justice: The Case for Polycentric Democracy.* New York, NY: Routledge.

Nguyen, C. Thi. 2019. "Games and the Art of Agency." *The Philosophical Review* 128 (4): 423–462.

Nguyen, C. Thi. 2020. *Games: Agency As Art.* Oxford: Oxford University Press.

Ostrom, Vincent, and Elinor Ostrom. 2002. "Public Goods and Public Choices." In *Polycentricity and Local Public Economies: Readings from the Workshop in Political Theory and Policy Analysis,* edited by Michael D. McGinns, 75–103. Ann Arbor, MI: University of Michigan Press.

Parfit, Derek. 2011. *On What Matters. Volumes 1 and 2.* Oxford: Oxford University Press.

Prichard, Harold A. 1912. "Does Moral Philosophy Rest on a Mistake?" *Mind* 21 (81): 21–37.

Rawls, John. 2001. *Justice as Fairness: A Restatement.* Edited by Erin Kelly. Cambridge, MA: The Belknap Press of Harvard University Press.

Rowe, Thomas. 2019. "Risk and the Unfairness of Some Being Better Off at the Expense of Others." *Journal of Ethics and Social Philosophy* 16 (1): 44–60.

Rowe, Thomas, and Alex Voorhoeve. 2018. "Egalitarianism under Severe Uncertainty." *Philosophy & Public Affairs* 46 (3): 239–268.

Scanlon, Thomas M. 1995. "Moral Theory: Understanding and Disagreement." *Philosophy and Phenomenological Research* 55 (2): 343–356.

Scanlon, Thomas M. 1998. *What We Owe to Each Other*. Cambridge, MA: The Belknap Press of Harvard University Press.

Scanlon, Thomas M., and Alex Voorhoeve. 2001. "Kant on the Cheap." *The Philosophers' Magazine* 16: 29–30.

Southwood, Nicholas. 2008. "A Deliberative Model of Contractualism." *Politics, Philosophy & Economics* 7 (2): 183–208.

Southwood, Nicholas. 2010. *Contractualism and the Foundations of Morality*. Oxford: Oxford University Press.

Southwood, Nicholas. 2019. "Contractualism for Us As We Are." *Philosophy and Phenomenological Research* 99 (3): 529–547.

Suikkanen, Jussi. 2019. "*Ex Ante* and *Ex Post* Contractualism: A Synthesis." *The Journal of Ethics* 23 (1): 77–98.

Valentini, Laura. 2021. "Respect for Persons and the Moral Force of Socially Constructed Norms." *Noûs* 55 (2): 385–408.

Van Schoelandt, Chad. 2019. "Between Traditional and Minimal Moralities." *Analysis* 79 (1): 128–140.

Van Schoelandt, Chad. 2020. "Functionalist Justice and Coordination." *Social Theory and Practice* 46 (2): 417–440.

Watson, Gary. 1998. "Some Considerations in Favor of Contractualism." In *Rational Commitment and Social Justice: Essays for Gregory Kavka*, edited by Jules L. Coleman and Christopher W. Morris, 168–185. Cambridge: Cambridge University Press.

Williams, Bernard. 1973. "A Critique of Utilitarianism." In *Utilitarianism: For and Against*, edited by John J. C. Smart and Bernard Williams, 77–150. Cambridge: Cambridge University Press.

Williams, Bernard. 1985. *Ethics and the Limits of Philosophy*. London: Fontana Press.

Michael Moehler is Founding Director and Core Faculty member of the Kellogg Center for Philosophy, Politics, and Economics at Virginia Tech. He received his Ph.D. in philosophy from the London School of Economics. His research and teaching expertise lie in the history of moral and political philosophy, rational choice theory, public reason theory, distributive justice, the welfare state, and political economy. He is author of *Minimal Morality: A Multilevel Social Contract Theory* (Oxford University Press) and *Contractarianism* (Cambridge University Press).
Contact e-mail: <moehler@vt.edu>