

Virtue, Creativity, and Ethos in Normative Political Thought

Marina Vahter

2026

Doctoral Dissertation

Defence: Dec. 17, 2026 at Tallinn University

Recommended citation: Vahter, M. (2026). *Virtue, creativity, and ethos in normative political thought* [Doctoral dissertation, Tallinn University].

© Marina Vahter 2026

TABLE OF CONTENTS

TABLE OF CONTENTS	2
PREFACE	5
ACKNOWLEDGMENTS	6
INTRODUCTION	7
1. DISTINGUISHING BETWEEN MORALITY AND ETHICS	14
1.1 Why is this distinction necessary?	14
1.2 Ethics and morals, ancient and modern	16
1.2.1 Shared roots, divergent paths.....	16
1.2.2 Modern times and the recovery of ethics.....	18
1.3 Nietzsche's genealogical critique of morality	21
1.4 Morality and ethics: working definitions and conceptions	25
1.4.1 The anatomy of the moral system	26
1.4.2 On ethical valuation	29
2. POLITICAL MORALISM	36
2.1 What is political moralism?	36
2.2 Rawls's conceptual moralism	38
2.2.1 Rawls on morality	38
2.2.2 Rawls's conceptual moralism in political liberalism.....	41
2.3 Justice	46
2.3.1 Justice and its intrinsic value	46
2.3.2 Why justice?	49
2.4 Politics.....	51
2.4.1 Politics as a fair system of cooperation	51
2.4.2 Liberal legitimacy	57
2.5 Political actors	62
2.5.1 Moral powers.....	62
2.5.2 Epistemic powers	63
2.5.3 Emotional powers, or the lack thereof.....	65
2.6 Critique of Rawls	68
2.6.1 (In)Feasibility	68
2.6.2 Power and the status quo.....	70
2.7 To conclude: does the political have to begin with morality?	74
2.7.1 Justice as the source of stability for the right reasons	75
2.7.2 Lessons from Rawls	79
3. POLITICAL REALISM	81
3.1 What is political realism?	81
3.2 Realism as anti-moralism	84
3.2.1 Williams: morality as a "peculiar institution"	85
3.2.2 Realist ideology critique.....	88
3.2.3 The wrong target?	93
3.3 The difficulties of excluding pre-political morality	94
3.3.1 Metaethical commitments of realism.....	94

3.3.2 The realist conception of the political	99
3.3.3 The realist search for purely political normativity	102
3.3.4 Realism and liberal legitimacy	105
3.4 Concluding remarks: political realism and its reductions	107
4. POLITICAL AGONISM	112
4.1 What is political agonism?	112
4.2 Bonnie Honig's joy of agonism	115
4.2.1 Virtù, not morals	116
4.2.2 How morals displace politics	118
4.2.3 Honig's agonism	123
4.3 William Connolly's agonistic respect	125
4.3.1 Connolly's descriptive and explanatory account of antagonism	125
4.3.2 Connolly's normative account of agonism	127
4.3.3 Critical responsiveness and justice	129
4.4 Chantal Mouffe's agonistic pluralism	132
4.4.1 Radical democracy and its open-endedness	133
4.4.2 Conflict	135
4.4.3 From antagonism to agonism	137
4.5 Morality and political agonism	142
4.6 To conclude: agonism and its discontents	144
5. VIRTUE AND ETHICS	150
5.1 On variations of virtue ethics	150
5.1.1 Aristotelian virtue ethics	150
5.1.2 Agent-based sentimentalism of Michael Slote	154
5.1.3 Pluralist and target-centred virtue ethics of Christine Swanton	157
5.2 Nietzsche and virtue ethics	160
5.2.1 The foundations of Nietzsche's ethics	160
5.2.2 Nietzsche's ethics of virtue	165
5.3 Contemporary attempts to bridge virtue ethics and political theory	171
5.4 To conclude: on the immoralism of virtues	174
6. THE ETHOS OF CREATIVE VALUATION	176
6.1 Preliminary outset: virtues, dispositions, and value	176
6.2 Working definitions: virtue, disposition, and good	177
6.2.1 Virtue as a capacity	178
6.2.2 Disposition	180
6.2.3 Good and bad virtues?	181
6.2.4 Social value	183
6.3 From virtue to disposition to social value	184
6.3.1 Virtues as epistemic capacities for valuation	184
6.3.2 Virtues and the process of valuation	186
6.3.3 From individual disposition to social virtue	191
6.4 Creativity as a virtue	193
6.5 What is virtue-ethical politics for?	197
6.5.1 Moralism, realism, agonism on politics and power	197
6.5.2 Power and politics: the virtue-ethical view	199
6.6 Virtue-ethical politics	202

6.6.1 Politics as constitutive resistance and individual empowerment.....	203
6.6.2 From self-correction to self-creation	205
6.6.3 Democracy as practice	207
6.6.4 Stability: weak and strong.....	211
6.6.5 From supervision and rules to considerations.....	214
6.7 Concluding remarks and possible objections	220
6.7.1 Core tenets of the ethos of valuation	220
6.7.2 Objections.....	224
CONCLUSION	233
REFERENCES	237

PREFACE

This study was spurred by a sense that political scholarship often treats values as something that stands in the way of doing ‘real politics’. That seemed to imply that questions of doing good and being good do not belong to politics, which I found puzzling. So I began to ask what it is about values that makes them appear to mess up the organisation of our common life, which in turn required asking what we mean when theorists and laypeople alike speak of foundational values in terms of morality and ethics. Surely it cannot be that we do not want to do good and be good in our common world – or can it?

Was I mistaken to think that every feature of social life is somehow connected to values? After all, social life is about human connections, and human connections are bound up with values. A cynic might say that social life is about survival. Yet survival, too, is ultimately a question of value – of who survives, why, and, most importantly, *how*. And what also seems true is that humans have created politics to somehow orchestrate this chaotic symphony of survival amid the vast differences in values. Because human societies are societies of culture, and culture is the kaleidoscopic animation of our imagination, which we forge into meaning through values, the very reason for doing politics becomes the exchange of values: values speaking to one another; values meeting, disagreeing, and creating together.

When people meet in the public sphere, what meets is their values. So values, in all their varieties, need to survive as well, because people are their values. This remains true also when human encounters are abstracted: when meetings among people are no longer direct but mediated by symbols, history, institutions, conventions, and the like. Politics is ever-present in both direct and indirect interactions because values are always present.

Yet it also seemed that, for many, politics is a matter of the survival of interests, and that political scholarship is uncomfortable with values because interests appear more tangible than the infinitely messy world of values. But interest-driven thinking, with its drive for certainty, turns politics into a matter of positionality: the question of how and why a particular position does or should prevail. Politics then becomes a social matrix that arranges the capacity to shape choices, while setting aside the exchange of the experience of valuing, creating, and making meaning.

I hope this study reflects my desire to bring the human being back into the realm of politics and treat politics as a meeting place of values, rather than a zero-sum game of interests and power.

ACKNOWLEDGMENTS

I am truly grateful to my reviewers, Kadri Simm and Tomas Kavaliauskas, for their valuable feedback and suggestions, and for asking thought-provoking questions. Without their input, this study would be considerably less coherent. I am also indebted to the members of the ECPR Standing Group on Methods of Normative Political Theory, with whom I have had the great pleasure of engaging over the years at conferences and workshops. These encounters have not only inspired parts of this study but have also broadened my horizons. Finally, I am grateful to the friends and colleagues who have taken an interest in this work and followed my journey through its various stages. Thank you all!

INTRODUCTION

*In the social world, we are brought together
by our basic need for recognition,
and torn apart by our failure to see pain.*

Nietzsche's Zarathustra suggests that what makes us human is our capacity to place values into things, and so to create meaning ([1883] 2006, On a Thousand and One Goals). Our commitments, beliefs, preferences, choices, and how we experience ourselves among others are all a matter of valuing; and valuing is in turn a matter of what meaning we create within these relations. Indeed, it is the process of making sense of others that animates the creativity within us and calls for its expression. When such expressions are brought together and exchanged, they become what we call culture. Whenever we create, we value, and in doing so we create meaning. The creation of meaning, then, is part of what it means to live.

This study is concerned with how humans, in virtue of their human traits, engage in the creative process of valuation. In particular, I am interested in ethical valuation, which I take to be at the core of human relations. Ethical valuation concerns the question of how we are, and might be, towards one another, whoever or whatever the other turns out to be. This is because ethical valuation concerns primarily the valuer rather than the object of valuation. By contrast, approaches that build on moral reasoning may ask 'what we owe to each other', as Scanlon (2000) does, by asking what is due between parties who hold one another mutually accountable. The study accordingly begins by drawing a substantive distinction between ethics and morality, as I take these two questions to lead in different directions when we think about how to organise our social life.

Because ethics is a social activity and entails an infinite variety of value conceptions, much of social life is necessarily bound up with politics, which attempts to orient our common life amid this diversity of values. I therefore bring together two activities of human life that cannot live with or without each other, philosophy of ethics and normative political theory, as the common life among the immense diversity of values has two dimensions: philosophically, it concerns the meaning of living together; and politically, it concerns how we organise this common life of ours.

Politics exists because we need to work out how to live together despite the different meanings we give to things. It is one way of expressing the human condition, bringing the creativity involved in 'figuring things out together' into the political agora. Throughout the history of philosophy, the role of values has

posed a persistent challenge to thinking about how we should live together politically. We cannot live without values, but because each person has their own lifeworld of values, living together as a collective and interdependent body is immensely complex. And this complexity is precisely what politics as an activity tackles, because with diversity comes disagreement, but common lives require some shared agreements. Politics is the agora for figuring out what agreements to make, and on which grounds.

Contemporary political philosophy has addressed the question of how to think about politics in the context of persistent value diversity in several ways, three of which I explore in this study. Rawls's (1996) political liberalism holds that a just society should be based on political principles that people with different worldviews can accept on their own terms. Political realism argues that politics has its own standards, grounded in order and legitimacy rather than in moral values that stand outside politics (Williams 2005; Geuss 2008). Political agonism thinks that consensus is neither attainable nor desirable, and that what disagreement needs is a form in which adversaries recognise one another's standing to disagree (Honig [1993] 2023; Connolly 1995; Mouffe 2013).

However, what looks like three answers to one question is actually three different questions about the level at which diversity-driven disagreement should be addressed. For Rawls, diversity is a problem of justification. Realists want to secure political order before justifying substantive commitments. Agonists view diversity-driven conflict as part of what politics is. For all their differences, the three converge on the notion that foundational commitments are too irreconcilable for politics to resolve, so politics is left to manage disagreement without engaging it at the level of substance.

I propose that there is a further level behind this foundational value diversity that has received little attention in political philosophy, namely the level at which values are formed. Value commitments and doctrines are essentially verdicts. But verdicts do not emerge from nowhere; they are not given, but are the result of a complex process of valuation. If we take verdicts to be incommensurable by looking at the verdicts themselves and not at how they were arrived at, we do not know what really makes them incommensurable, or whether they really are so, because we do not know the reasons behind them. This is what I call the priority of value over valuation, and it is the common feature I find in moralism, realism and agonism throughout the chapters that follow. It follows that the diversity at issue in politics is not, in the first instance, a diversity of commitments but a diversity of reasons.

Because values are an integral part of life, and of political theorising too, engagement with them is inevitable, whether by affirming or by rejecting them.

Each theory about political arrangements ought therefore first to understand its own value commitments clearly. In other words, every political arrangement has an ethos, and making that ethos explicit can improve understanding between different perspectives on politics.

This is not to say that the approaches examined here lack an ethos altogether. Rawlsian moralism carries an ethos of reasonability, realism's is anti-utopian (Geuss 2008, 24-25), and agonism is marked by an ethos of contingency (Selg 2012). I argue, rather, that these approaches leave their ethos insufficiently acknowledged, so that commitments doing normative work are not themselves examined as commitments. This is, for example, why I argue that realism takes stability and legitimacy as standards "without explicitly recognising them as values" (3.4).

So, ethos is understood here as a normative background orientation that reflects the values and ethical commitments of a framework. If ethical thought is an inalienable part of human life, then any serious normative argument needs to make explicit the value conceptions on which it relies and which it promotes and, more importantly, why those values matter in the particular context. Understanding the ethos matters because it orients the reasons for which we prioritise certain actions over others. And because values do not emerge from nowhere, nor do they carry a uniform meaning, we need to dig deeper into how values are shaped. Rawls claims, for instance, that a society functions only insofar as it is regarded as just by its members, on the basis that what people share is a sense of justice, while disagreeing about its content (1996, 391). It is not sufficient, then, to understand what each of them means by 'just'; we need to understand how that meaning is formed.

Thus, insofar as values are the result of creating meaning for phenomena, we need to understand the process of valuation. This study therefore brings the ethos of theorising to the forefront and shows how theorising at that level might be done. In particular, I sketch an ethos of creative valuation which roots valuation in virtues, understood as the human trait that carries the capacity to value. In the traditional sense of virtue ethics, virtues enable responding well to antecedently established values. On my account, virtues are epistemic capacities through which values acquire meaning. To give a deliberately oversimplified example, the value of social trust differs depending on whether the virtues of sincerity or accuracy are involved. This example, drawn from Williams's account of truthfulness (2002), runs through chapter 6. Of course, the world of meaning-making is infinitely complex; virtues do not have fixed content, and they are always employed in configuration, not alone. Virtue, on my account, designates only a rough approximation of the kind of meaning it can bring to a value. This

too is deliberate, because I do not think we should ask what a word could possibly entail, but should focus instead on the process and the phenomenon in the social world, and simply use language to paint a certain picture of the world as we see it, considering that words and virtues alike acquire meaning upon being exercised. Accordingly, my proposal shifts how we think about values from prioritising value to prioritising the process of valuation.

Valuation is understood here not as evaluating something as good or bad, but as a complex process of meaning-making which starts from attention to relevance. It does not yield a verdict as moral judgement does, and it is not directed at decision-making. It is relational, reflexive, affectively charged and internally motivated, and it proceeds in a narrative rather than an inferential mode. Nor does the refusal of verdicts make it merely descriptive – to hold that values are created and re-created by those who live by them is already a claim about how we ought to engage with the values by which we live.

From this I sketch a democratic ethos that I have termed self-creative democracy (6.6). Its two grounding ideas are constitutive resistance – the engageable difference that makes valuation possible in the first place, and which must be roughly equal if it is to be engaged rather than eliminated (6.5) – and the framing of political arrangements in terms of considerations rather than rules, since a rule answers a question in advance, whereas a consideration says what is to be attended to and leaves the answer open (6.6.5). To be sure, the ethos offers no institutional model, let alone particular policies; that will have to be left for future studies. My aim here is simply to show why an ethos of theorising is needed and what one can look like.

To this end, the study draws a sharp distinction between ethical and moral thought in order to foreground virtue ethics, drawing inspiration primarily from Nietzsche. Ethics is viewed as an epistemic enterprise concerned with how values are generated, whereas morality is concerned with determinate judgements of right and wrong that prescribe action. In this respect, while moralism and realism tend to portray political subjects as largely reactive to social conditions, and agonism presents them as active agents, creativity-centred virtue ethics goes further by depicting political subjects as proactive agents who not only respond to tensions but also anticipate and even generate friction in creative and potentially empowering ways.

The contribution of this work is fivefold. First, it makes a *metatheoretical* contribution to political philosophy by arguing that every normative political theory has an ethos. Theories therefore need to examine and make explicit their own ethos, that is, the values and assumptions that guide them. Second, it makes a *diagnostic* contribution. Moralism, realism, and agonism consider fundamental

disagreement about values ultimately irreconcilable because they approach it as a disagreement over the solution rather than over how the underlying problem is conceptualised. On the view advanced in this study, for example, disagreement over whether assisted suicide is permissible may indeed be irreconcilable at the level of the verdict, but the reasons behind that verdict may be the same – and, as such, reconcilable. Third, it makes a *metatheoretical contribution to the philosophy of ethics*. Understanding the reasons behind a value verdict requires an account of the capacities through which values acquire meaning. This is what the reconstruction of virtues as epistemic capacities is intended to provide. It understands virtues as epistemic capacities through which social values acquire meaning, rather than as dispositions for responding well to values that are already fixed. Fourth, it makes a *positive contribution to the philosophy of ethics* by developing a bridge between Nietzsche’s themes of creativity, the will to power, *amor fati*, and struggle, and Iris Murdoch’s concept of attention, while advancing the priority of valuation over value. This provides the basis for understanding what I call constitutive resistance in ethical valuation. Fifth, it contributes to *political philosophy* by outlining an ethos of self-creative democracy. It shows how a commitment to creative valuation, which does not seek to fix the meaning of values, can be applied to democratic politics through three concepts without requiring agreement on foundational values from citizens: constitutive resistance, considerations rather than rules, and the distinction between coordinative and supervisory order.

This study is structured in six chapters. The first chapter sets the scene by discussing the concepts of morality and ethics, first tracing their historical development and usage in political philosophy, and then sketching how morality and ethics are conceptualised in this study. It sets out the anatomy of the moral system that the later chapters take as their target (1.4.1), and the conception of ethical valuation that the proposal rests on: relevance rather than verdict, sense and sensibility as its two cognitive dimensions, and Murdoch’s example of M&D as the case to which chapters 5 and 6 return (1.4.2).

The second chapter explores Rawls’s political liberalism as an example of political moralism. It outlines the central features of justice-oriented political thought; what we learn from Rawls is that social relations are deeply entwined with the conception that they ought to be just. The chapter’s central interpretive claim is that Rawls’s moralism is conceptual rather than substantive: morality supplies the framework within which evaluative questions can be asked, but it does not provide the substantive answers, i.e., what is just (2.2.2). The chapter also develops the idea of moral truth as *unowned* (2.4.2) and reads Rawlsian idealisation as heuristic rather than telic (2.6.1), which is how the feasibility objection is answered.

The third chapter examines the political realist rejection of pre-political morality, which instead grounds normativity in political realities such as power, legitimacy and practical reasoning, on the argument that morality is too contested to provide a stable foundation for politics. I examine the four candidates realists have offered for a purely political normativity and argue that none is self-standing (3.3.3), and that legitimacy cannot supply the standard by which its own corruption can be detected (3.3.4). Realist ideology critique nevertheless locates morality in the right place, among the vocabularies through which power justifies itself (3.2.2).

The fourth chapter examines agonism, which, like realism, rejects pre-political morality; unlike realism, however, it does not treat legitimate stability as the aim of politics. Instead, it understands democratic conflict to enable marginalised voices to challenge dominant norms and power structures, preventing politics from collapsing into managerial administration or moralised closure, while avoiding framing politics in ethical terms. The chapter focuses on three thinkers. Bonnie Honig argues that agonistic disruption must be actively cultivated, as stabilisation is the default tendency of social life. William Connolly locates agonism in the dispositions of individuals, emphasising critical responsiveness and agonistic respect in the context of ontological pluralism. Chantal Mouffe, by contrast, advances a hegemonic account centred on collective mobilisation and the construction of a shared civic framework. Agonism comes closest of the three to escaping the priority of value over valuation, but the framework within which contestation takes place still rests on values settled in advance, and the accounts do not explain how the dispositions they require of political actors are formed, as they treat those dispositions as outcomes of the very practices they are meant to make possible (4.6), and this is the gap chapter 6 sets out to fill.

The fifth chapter situates virtue within ethical reasoning, discussing Aristotle's account, Michael Slote's agent-based sentimentalist account, and Christine Swanton's target-centred pluralist framework, including a critical examination of attempts to render virtues action-guiding. Finally, it explores Nietzsche's virtue ethics. The three mainstream accounts agree that the good is specified independently of the exercise of virtue, and the exercise is then assessed against it, while disagreeing over what virtues do: Aristotle locates the good in the activity, Slote in the motive, Swanton in the target (5.4). The chapter also surveys existing attempts to bridge virtue ethics and political theory and identifies the pattern they risk falling into, terming it 'virtue moralism': virtues pressed into the service of universal, obligation-generating requirements (5.3).

The sixth and final chapter draws on Nietzsche to develop an ethos of virtue-ethical creative valuation. It offers a path for thinking about the ethos underlying

normative theorising, but it also advances a particular ethos that foregrounds the process of valuation as an open-ended endeavour and as the means of engaging creatively with politics. The ethos treats resistance as the source of creative joy and empowerment, allowing the meaning of social value to emerge. The chapter starts by outlining how I conceptualise the process of valuation, and how its different features – attention, virtue, disposition – create the meaning of social values. It then focuses on the concept of creativity and how it functions as a virtue, before sketching the outlines of an ethos of a self-creative democratic order. Along the way it distinguishes supervisory from coordinative order, arguing that monitoring undermines the very conditions under which the goods it is meant to secure, trust above all, are created, and it argues that political arrangements are better framed as considerations than as rules (6.6.5). In the last section (6.7), I summarise the core tenets of the proposed ethos and set out possible objections with replies.

1. DISTINGUISHING BETWEEN MORALITY AND ETHICS

1.1 WHY IS THIS DISTINCTION NECESSARY?

In everyday use, and often in scholarly writing as well, morality and ethics are employed interchangeably, referring broadly to something bound up with foundational values. Yet these terms can be understood in many ways, and their content therefore requires careful delineation. This is all the more relevant for this study, since the creativity-centred virtue-ethical proposal developed in chapter 6 distances itself sharply from moral reasoning and adopts a particular view of ethical valuation, one that departs in certain respects from the current mainstream.

That said, the critical target of morality in this study is narrow. The critique of morality advanced in the chapters that follow is directed at a historically specific formation of practical thought – principle-centred, universalist, obligation-generating, and oriented towards determinate closure – an understanding broadly shared by a line of thinkers from Nietzsche through Elizabeth Anscombe to Bernard Williams. Nietzsche offers a genealogical-psychological account of *herd morality* as a formation with a distinct social function ([1887] 2007, §§1.10-1.11, see also 1.3); Anscombe conceptualises morality as a “law conception of ethics”, which we inherited from the earlier Christian conception of ethics as obligation (1958, 5-6); and Williams treats it as a structure – “a peculiar institution” organised around obligation (1985, 174-196, see also 3.2.1).

Furthermore, modern philosophy offers no settled usage of these terms to rely upon. Section 1.2 traces in detail how the two have divergent histories, and how they are today divided between two broad conventions, corresponding, albeit somewhat loosely, to the analytic and continental traditions of philosophy. On the first, morality is a first-order code of conduct and ethics its philosophical study. On the second, ethics designates the broad question of how to think about living well, while morality is a narrower system of rules. The second approach is also where this work situates itself.

Distinguishing between morality and ethics also brings us closer to understanding what I take to be a source of the apparent impasse between moralism, realism, and agonism – an impasse that is, I argue, partly an artefact of considering values solely through the lens of the principle- and obligation-centred system of morality. The realist Geuss goes so far as to describe Rawlsian (moralist) political liberalism as an “ethics-first” approach, a form of “applied

ethics” (2008, 1). Yet realists and agonists alike prove unable to entirely avoid pre-political guidance of their own in their attempts to construct a purely political legitimacy (Leader Maynard and Worsnip 2018; Erman and Möller 2015), and agonists cannot do without the intrinsic worth of their pluralist contestation (Fossen 2008, 378). Thus, in equating the ethical with the moral, realists and agonists risk throwing out the baby with the bathwater. Rawls, for his part, speaks of rights and obligations, arranged into principles in pursuit of consensus in the public sphere, and defines “moral theory” as the second-order (meta-level) study of moral structures (1974, 6), while leaving the question of how to live to the private sphere (1996, 175) – to what he calls the “background culture” (ibid., 220). So, after discussing moralism, realism, and agonism in chapters 2–4, I will argue that once morality and ethics are distinguished, the impasse loses much of its air of necessity, since what realists and agonists rightly reject is the morality system, and what they cannot do without is ethics.

Finally, the creativity-centred, virtue-ethical approach to normative theorising proposed in chapter 6 – a non-prescriptive, non-obligation-generating mode of normative thought – is intelligible only once moral reasoning and ethical valuation are held apart. To emphasise, my aim is not to establish what the terms morality and ethics essentially mean. Rather, it is to identify certain phenomena in the world and to clarify how I use these concepts to describe them, thereby avoiding confusion with other uses of the same terms. The more specific definitions and conceptions employed in this study are set out in section 1.4, but it should be kept in mind that these terms are intended to represent phenomena, rather than to prescribe what the words mean (see also Toomela 2016, 92 fn. 1). In short, when I use these concepts to sketch out my proposal, I do not intend to structure the meaning of words, but to use words to structure meaning.

To understand the rationale behind the particular conceptions of morality and ethics employed here, it is useful first to trace the development of these concepts and their modern uses (1.2), then to focus specifically on Nietzsche’s genealogical critique of morality (1.3), as it underlies the conception of morality employed in this study, and finally to provide working definitions of morality and ethical valuation (1.4). Section 1.4.1 presents a synthesis of accounts of modern morality intended to specify the object of my critique in this study, while section 1.4.2 sets out what I mean by ethics and ethical valuation in the context of the virtue-ethical approach to theorising developed in chapter 6. However, when I discuss morality or ethics in relation to moralism, realism, and agonism, I do so from within their respective conceptions of these terms, unless indicated otherwise.

1.2 ETHICS AND MORALS, ANCIENT AND MODERN

1.2.1 Shared roots, divergent paths

The ambiguity around the meaning of ethics and morals can be explained by their shared roots. In Ancient Greece, ethical thought did not distinguish between what we today call moral and non-moral thinking (Annas 1992, 120). The modern use inherits the name of the field through Aristotle, who used *ēthikē* for the virtue of character and derived it from habituation, saying that virtue of character “results from habit (*ethos*) – indeed, this is the source of the name *ēthikē*, which derives with a minor variation from *ethos*” (NE II 1.1103a17-18). Put otherwise, Aristotelian ethics is an inquiry into how one comes to be the person one might become.

The term “morality” comes from Cicero, who coined *moralis* from *mores* (the plural of *mos*, meaning custom, habit, or convention) to translate the Greek *ēthikos* (De Fato I.1). *Moralis* is generally understood to have been intended to carry the same sense of character as the Greek *ēthos* (MacIntyre [1981] 2007, 38). However, this interpretation of Cicero is not universally accepted. Gordon (2013), for example, argues that *moralis* refers only to convention in a general sense and does not include the idea of character. He argues that this narrower meaning was necessary because a convention can itself be immoral, so morality cannot simply mean following convention. My purpose in discussing these origins, however, is not to determine which interpretation is correct. Rather, I want to show that the terminology has never had a single, settled meaning.

Thus, in the broadest terms, Ancient Greek and Roman ethics and morals were likely to point towards character and its formation, and only later, as Williams notes, did the Latin branch, developing in particular “a special notion of obligation” (1985, 6), come to point towards the meaning of social custom and its codification. Let us now briefly examine how this happened.

Ancient ethical reflection was organised around Socrates’s question of “how should one live?” (Williams 1985, 4), a question concerning the whole of a life, not particular deeds. The central evaluative concepts were *aretaic*, i.e., virtue understood as excellence, where virtue is not a moral property of the person but the excellent performance of a function, just as a knife performs its function well when it cuts well. Most importantly, the domain with which *aretaic* ethics is concerned is the individual, and where deontic notions of obligation, law, and guilt do appear (notions we take to belong to modern morality), they function as local tools, without universal ambition (Annas 1992, 121). To reiterate, ancient ethics is not concerned with other-regardingness as a means of social coordination, in which the right (action) answers to some universal point of

reference, whether because there is broad agreement that it is right, or because the good is derivable by calculation, or because it corresponds to a general principle. This does not mean that ancient ethics was removed from social relations; justice, for instance, was considered one of the virtues (excellences). But it was a virtue because it was valuable for its own sake in the context of the good life of 'one', not of 'every single one' or of 'everyone' (see also Williams 1985, 4–5).

Thus, since morality in the modern sense did not exist in ancient ethics, it had to be invented. Tracing its emergence leads us to the revaluation of values and its Christian inheritance. Following Nietzsche, we can say that the roots of morality lie in the reactive emergence of slave morality, which begins with the resentment of those whose weakness prevents them from living the self-affirming life of the masters ([1887] 2007, §§I.7, I.10). Anscombe offers a similar account, although hers is conceptual rather than genealogical. She argues that the "law conception of ethics" (1958, 5, original emphasis) transformed the question of how to live into the question of what is required, forbidden, and permitted by a divine legislator. On this view, our concepts of moral obligation are remnants of a framework human society has abandoned (1958, 5–6). So, what Nietzsche and Anscombe describe is the emergence of a particular form of valuation. The later codification of this form of valuation as law is a separate development, generally taken to begin with Aquinas, whose natural law theory was the first to systematise morality in terms of law (Schneewind 1998, 19–21).

It was not until the Enlightenment and Kant's *Moralität*, however, that morality acquired the full architecture of its modern form, together with an intellectual autonomy of its own. Only then, amid the growing dominance of rational method, did philosophers begin to seek a properly rational justification for moral codification (MacIntyre [1981] 2007, 39). The search for regularities and rational order that characterised the scientific ethos of the Enlightenment thus extended into the human and social spheres. Whereas ancient philosophy did not sharply distinguish between the physical, social, and human domains, the Enlightenment transformed these into autonomous disciplines.

Kant's *Moralität* establishes a system of universally valid principles binding on all rational agents. The categorical imperative commands that one act only on maxims that one could will as universal law, admitting no exception on the grounds of circumstance (Kant [1785] 2008, G4:421). Kant's system therefore incorporates every feature of modern morality described in section 1.4: principle-centredness, universality, obligatoriness, and the aspiration to determinate closure as the assumption that, for every moral dilemma, practical reason can provide a correct answer.

Although Kant's doctrine became widely adopted, it was also criticised. For example, MacIntyre observes that Hegel traced the assumptions underlying Kant's ethics to Protestant thought, arguing that Kant secularised them and that they later became a foundation of liberalism (MacIntyre [1981] 2007, 266). But Hegel ([1821] 1991) was also the first to clearly distinguish between Kantian *Moralität* and ethical life (*Sittlichkeit*). For Hegel, *Sittlichkeit* refers both to the concrete institutions of a community and to an individual's lived identification with those institutions (Wood, in Hegel [1821] 1991, 404), whereas *Moralität* is the subjective will reflected into itself, legislating solely from its own certainty. Hegel argues that, because the categorical imperative lacks particular content, it collapses into "an empty formalism" ([1821] 1991, §135R), and *Sittlichkeit*, in turn, fills this void with the substantive content provided by the institutionalised customs and practices of a community. In short, Hegel conceives morality as a narrow mode of ethical thought that mistakenly attempts, through the universalising categorical imperative, to govern the whole.

1.2.2 Modern times and the recovery of ethics

In the wake of Kant, moral philosophy from the turn of the twentieth century was adopted by the broader analytic turn, and ethics accordingly became 'analytic ethics'¹ in both its normative and applied strands. In normative moral philosophy, morality became understood as a first-order code of conduct, a set of norms governing action whose proper method was one of structural logic, form, and conceptual coherence, and ethics came to designate the study of moral codification, most prominently through metaethics.

Metaethics is indeed an overwhelmingly analytic enterprise, so much so that the Internet Encyclopedia of Philosophy opens its entry by stating plainly that metaethics is a branch of analytic philosophy concerned with the "status, foundations, and scope of moral values, properties, and words" (Delapp, n.d.). We can trace this to G. E. Moore's separation of the question of whether a particular thing is good from the question of what "good" means as a concept ([1903] 2002, §6). The metaethical positions that followed – cognitivism, non-cognitivism, realism, anti-realism, emotivism, error theory, naturalism, intuitionism – form a distinctly analytic apparatus, concerned with the logical form and truth-aptness of moral claims (Sayre-McCord 2023).

¹ The divide that is relevant here is methodological, between the formal method, which is concerned with the logical structure and coherence of moral judgement, and the genealogical method, which is concerned with how moral meaning came about. These methods are roughly represented by the analytic and continental traditions respectively, although this division is somewhat uneven. For example, Anscombe and Williams, both discussed in this study, were formed within Anglophone philosophy, but both worked against its formalism. So, wherever I use the analytic/continental labels, they should be taken only as a rough guide to the traditions in which each method has predominated.

This, in turn, has led formal methods to conceive of morality as a set of principles that ought, first and foremost, to conform to the logical structure and coherence of judgement. In its focus on form, this method treats the question of value as a technical matter of how a proposition sustains itself. Put simply, it deals with the configuration of morality needed to distinguish right action from wrong, placing the validity of its arguments wholly in the hands of logic and coherence.

But form might say very little, if anything at all, about people's moral or ethical motivation, whether that motivation is understood as subjective or objective, or about how they come to see a situation as calling for a value-driven response in the first place. Even if people do reason about other-regardingness and right action in the terms the formal method supplies, it is far from clear that such reasoning is what motivates action (Rosati 2016). Something like this is also suggested in moral psychology. Jonathan Haidt's (2001) social intuitionist model suggests that moral judgements begin with quick, emotionally charged intuitions, and that moral decisions are justified post hoc through cognitive reasoning. In the studies Haidt discusses, participants condemned situations they perceived as violating moral norms, even though those situations caused no real harm, and even after they could not explain exactly why they thought those situations were wrong (ibid., 817). So, explaining where our judgements come from can expose an assumption the formal method has, namely that people will act on reasons when they see them. But the formal method does not ask how someone becomes the kind of person who can see those reasons, let alone be moved by them.

Relatedly, the moral method presupposes a particular standard of judgement and the rationalist capacities it requires. Rawls's method of reflective equilibrium, for instance, presupposes the moral and intellectual capacities of rational intuitionism without asking how those capacities are themselves formed, and, more importantly, without admitting emotive capacities, which are a principal wellspring of motivation (see 2.5.3). While striving for internal coherence, then, formal moral philosophy tends to drain the nuance from the human condition, reducing it to predictable machinery. It thereby sets aside the possibility that the incoherence and error arising from emotive motivational force ought not to be divorced from the "rightness" of an action, if for no other reason than that error and moral learning are inseparable (Tangney et al. 2007).

Further, the governing norm of internal coherence also heightens the risk of self-justification, since any challenge mounted from outside analytic ethics' own terms comes to look like a category mistake. Analytic ethics, in this respect an instance of what Williams calls the morality system's "deeply rooted and still powerful misconception of life" (1985, 218), risks locking itself into a comfortable mode of justification that asks only whether a commitment meets existing

standards, rather than how those standards, and the commitments judged by them, came to be constituted within their context. That is not to say, however, that analytic ethicists are necessarily seeking to establish the ‘one right’ morality. This is a richly diverse field of positions, united only by the conviction that formal analysis is the best way to approach ethical concepts. Its inquiry accordingly remains at the level of the concept itself, since questions such as ‘how should one live?’ cannot be answered by formal analysis alone, pertaining as they do to meaning-making, arguably a far more creative endeavour than coherence-seeking allows.

Continental philosophy, by contrast, is well equipped to consider ethics as the question of how to live, or of whom to become. It is freer in form and reaches where formal analysis cannot, because it asks different questions, being interested not only in making sense of the human condition in rational terms but also in animating it. Analytic metaethics asks how moral concepts are organised and what they logically entail; the continental tradition asks what they come to mean for those who live by them, approaching this existentially, phenomenologically, or genealogically. Kierkegaard, Nietzsche, and Sartre each insist, to name only a few, that a life cannot be made right by following principles correctly, while Lévinas (1996) proposes that ethics is the starting point of philosophy as such, prior to ontology or metaphysics, locating it in the event of encountering “the Other”.

When it comes to morality, continental philosophy treats it as a socio-historically contingent system of norms and asks how that system, and its meaning, came about. The most prominent example is Nietzsche’s genealogical inquiry into the origins of moral meaning and practice, which interrogates the conditions of moral valuation and insists that moral values are themselves something to be critiqued, that is, that the value of those values must itself be called into question ([1887] 2007, Preface §6). Continental ethics does not, then, construct moral theories in order to establish frameworks for valuation; it looks for ways to critique them. Section 1.3 offers a more detailed overview of Nietzsche’s genealogical account.

The virtue-ethical turn initiated after the Second World War by Elizabeth Anscombe (1958) returned to the aretaic vocabulary of ancient Greece, in revolt against the logical positivism that had dominated before the war and that Ayer’s expressivism exemplified. The war itself sharpened the importance of the substantive content of values, prompting a move from treating morality as a kind of speech act towards grounding it in the metaphysical. Anscombe’s call became central to the development of neo-Aristotelianism, which, like continental philosophy, treats morality as a conceptual fossil, what Anscombe saw as the grammar of obligation without a legislator (*ibid.*, 1). Its exponents cannot be

assigned to the continental tradition, having been trained in the analytic one, though little turns on that here.

Ancient and modern virtue ethics are examined in greater detail in chapter 5. Here, in a discussion of the shared roots and subsequent divergence of ethical thought, the story would be incomplete were we to claim that moral thought and ethical thought parted ways for good. Philippa Foot (2001) is a good example of how moral principles and ethical reasoning can be reconciled. She was Anscombe's peer at Oxford, alongside Mary Midgley and Iris Murdoch (Mac Cumhaill and Wiseman 2022).

Foot is credited, alongside Anscombe, with reviving virtue in philosophy, on the strength of a conviction that runs throughout her work: that virtues and vices are more fundamental notions than rights, duties, or consequences (2002). She herself, however, explicitly disavowed any allegiance to virtue ethics (2004, 2), a disavowal attributable to her particular reading of modern virtue ethics amid the diversity of the field. Her theory is better considered not a rival to consequentialism and deontology but, as she herself described it, a stand against metaethical non-cognitivism (2001, 5). What Foot defends is an ethical naturalism that roots morality in human nature, of which the virtues are a part. Drawing on Aristotle's teleological account of the species, according to which seeking the good is the function of human beings (ibid., 33), her "natural norms" (ibid., 25–37) serve as the standard by which a human being is judged a good member of the kind. Foot's ethical naturalism might thus be characterised not as 'virtuist' but as ethical in a broader sense, since she rejects the Kantian claim that principles bind human beings simply by virtue of their rationality, while affirming that human beings share a common nature. Her central question is what makes a person good as a human being, and although her position is still objectivist, it focuses on the human being rather than, as much of modern Enlightenment moral thought does, on human actions.

There is, of course, a notable diversity among virtue ethicists, some of which is discussed in chapter 5. What they share, however, is a rejection of the Kantian and consequentialist view of morality as a separate system of universal rules that applies to rational agents without regard to human nature or history. For virtue ethicists, ethics is instead about human character and what it means to live a good life.

1.3 NIETZSCHE'S GENEALOGICAL CRITIQUE OF MORALITY

Nietzsche's *On the Genealogy of Morality* ([1887] 2007) is one of the earliest and most systematic critiques of where modern morality came from and how it developed.

He examines the historical conditions that shaped modern moral thinking, especially the Enlightenment's emphasis on rationality and Kantian ideas of reason and duty. These ideas became especially influential in the nineteenth century and continue to shape moral thought today. Although earlier thinkers such as Kierkegaard and Herder had already challenged Enlightenment moral philosophy (Weston 1994; Angier 2016; Piirimäe 2020), their interventions were neither as extensive nor as influential as Nietzsche's. This study accordingly draws substantially, though not exclusively, on his account of the human condition in developing a virtue-ethical framework for political theorising.

Nietzsche traces the emergence and development of moral values to show that the Enlightenment moral rationality of concepts such as 'good' and 'evil' is rooted in the "slave revolt in morality" ([1886] 1998, §260), the reactive inversion of an earlier master morality carried through the Judeo-Christian tradition ([1887] 2007, §§I.1-I.17). Master morality, associated with social elites, is defined by strength, excellence, and an active, notably unreflective orientation towards life. 'Masters' exercise power and possess the freedom to live as they wish. Because this dominance made it difficult for those lacking such strength to resist 'master' values and pursue a meaningful existence on their own terms, slave morality developed as a form of revenge. But this revenge did not take the form of direct aggression; instead it worked by inverting the values of master morality to construct an alternative moral foundation: pride became humility, boldness became patience, strength and self-affirmation became something 'evil', while self-denial became honourable. In this way, the resentment characteristic of slave morality transformed weakness into moral superiority, redefining virtue.

Within the social realm, individual slave morality evolved into the "herd morality" ([1886] 1998, §199), which Nietzsche sees as characteristic of modern European moral life. From this collective herd morality arises the concept of duty, understood as responsibility towards others, which becomes the foundation of a morality claiming validity for everyone. Here, the sense of obligation turns resentment inward, leading to feelings of guilt that arise from treating selflessness as an ideal. For Nietzsche, guilt (*Schuld*) is not the recognition of one's wrongdoing but the internalised sense of debt (*Schulden*) owed to an authority (Hatab 2008, 85).

By making self-denial a moral ideal, morality encourages people to live together by conforming to a submissive way of life. It also gives personal responsibility a clear meaning by offering a particular vision of the future, but this vision assumes that human goals and social and psychological conditions can remain fixed and stable (Ansell-Pearson 2007, xxvi). Because neither the individual 'slave' nor the social 'herd', shaped by resentment and dependence, can affirm life on their

own terms, they need to place their trust in an external authority, an architect who claims to provide 'objective' truths. So, Nietzsche's argument tells us that there is no objective reason to follow moral rules. What modern Western culture calls morality is instead a historical product of resentment, rather than a reflection of divine or rational truth. Similarly, master morality, which is based on strength, is simply an expression of the will to power of particular groups (ibid., xix).

A further question, which Donovan Miyasaki takes up, is how the people who create these moral values emerge in the first place, that is, the people who decide what counts as morally good (Miyasaki 2022a, 222–232). Miyasaki (ibid., 228) points out that Nietzsche's masters occupy their position of nobility by accident of birth, being simply the "well-born" ([1887] 2007, §1.10), which means that the producers of the meaning of 'good' are themselves the product of social structures that generate divisions and inequalities, sorting a population into the powerful and the powerless. This should not be read, as Miyasaki cautions, as making morality wholly a creation of the social system, with no room for individual agency. The point is that social structures make it easier for people who already have material resources, or the 'well-born', to gain even more, and subsequently, these structures reinforce the inequality that helped create their position in the first place. Ideational resources, such as education, aptitudes, and character traits, are distributed by the same mechanism, following individuals' positions within the structure. Understanding morality, or the process of moralisation, thus requires analysing how these conditions shape the 'slaves' in response to the power of the 'masters', through the continual reproduction of "the constant feeling of inferiority, a consistent and stable psychological state grounded in stable political, and ultimately material, conditions" (Miyasaki 2022a, 225).

The emphasis on social conditions, rather than on social mechanisms such as power, indicates that observing morality's emergence from social hierarchy and inequality does not, in itself, imply that Nietzsche endorses strict egalitarianism, nor is he accurately read as a straightforward authoritarian. In other words, it does not follow that the hierarchy from which morality arises ought to be corrected, or that morality itself somehow ought to be rescued from these structures. This is not the paradigm within which Nietzsche operates at all. His non-foundational openness, expressed through agonism, perspectivism, and scepticism, conflicts instead with the closure that authoritarian politics requires, and some scholars indeed find in his philosophy the potential for interpretations that lean left rather than authoritarian right (Miyasaki 2022b). What his critique targets is "egalitarian mediocrity, normalizing conformity, and a generalized model of rational agency – all of which are taken to subvert the development of cultural excellence, particularly those rare creators who are essential to the

furtherance of life” (Hatab 2000, 39). Put otherwise, on Hatab’s reading, Nietzsche suggests that when equality becomes an end in itself, society leans towards settling for the average, since normalising conformity actively manufactures sameness through education, morality, and public opinion, thereby removing the aspiration to excel.

The emphasis on excellence invites the objection that Nietzsche’s account is elitist and therefore anti-democratic, as his apparent admiration for the aristocratic, resistance-seeking ethics of the pre-Christian Greeks seems at odds with the emancipatory commitments that democracy is generally understood to involve. A possible response would be to argue that, for Nietzsche, resistance must be of roughly equal strength to be meaningful. An opponent who is too weak offers little to overcome, while opposition that is too strong makes meaningful competition impossible. Neither can create the kind of engagement that the agon is meant to foster. Thus, the aristocratic ideal is not about domination but about having a worthy adversary in someone who can offer meaningful resistance, creating a “pleasure in the other’s equality as resistance” (Miyasaki 2022a, 91). Rough equality is therefore not an obstacle to living well, but a condition for it. On this view, democracy is not sustained by settling political conflicts once and for all, but by continually creating and re-creating itself through those conflicts. Seen this way, Nietzsche’s view is much less hostile to democratic life than it may first appear, although the question of elitism is discussed further in section 6.5.

Nietzsche’s perspective is thus particularly informative for understanding contemporary approaches to morality among moralists, realists, and agonists. His genealogical method has directly influenced the political realism of Geuss (2008) and Williams (2002), as well as the agonistic thought of Connolly ([1993] 2002, 2005) and Honig ([1993] 2023), while also provoking responses from neo-Aristotelian thinkers including Rosalind Hursthouse (1999) and Philippa Foot (2001, 2002). Rawls’s engagement with Nietzsche, by contrast, positions Nietzsche as a critic of the moral foundations of egalitarian justice, since his creativity-oriented, perfectionist outlook rejects the very morality of equality underpinning justice as fairness. Accordingly, chapters 2–4 examine how moralists, realists, and agonists conceive of morality, providing a foundation for the subsequent exploration of a virtue-ethical alternative to these three perspectives (chapter 6).

To summarise, Nietzsche’s genealogical account of morality should not be read as an epistemic critique. As Srinivasan notes, he is indifferent to whether moral beliefs are true or moral concepts apt (2019, 141), attributing such fascination with error to the “English psychologists” from whom he sharply distinguishes himself (Nietzsche [1887] 2007, §§I.1-I.3). *The Genealogy*, on Srinivasan’s reading, is better understood as asking “what our representations do [...] and what we might do

with them” (2019, 141), that is, as an inquiry into their function: “What practices and forms of life do they help sustain, what sort of person do they help construct, and whose power do they help entrench?” (ibid., 142). Accordingly, Nietzsche’s critique is aimed at the Christian moral tradition as it was transformed into the Kantian and consequentialist morality, which inherited from Christian morality the notion of obligation and universal rules. What Nietzsche exposes is not that these claims are false, the product of some defective epistemic practice, but that they function quite differently from how they advertise themselves.

1.4 MORALITY AND ETHICS: WORKING DEFINITIONS AND CONCEPTIONS

The previous sections showed that morality and ethics have a wide range of meanings, and while I have surveyed only a small fraction of that variation, I hope it has been enough to establish that no settled account exists of any of the concepts bound up with valuation, and therefore also with evaluation. I take it as a working assumption, if not an outright conviction, not only that they will never be settled, but that they should always be held open. My reason is not merely that agreement is unlikely; I also argue that settling the meaning of a value forecloses the valuation through which it acquires meaning in the first place, as I will discuss in chapter 6. As this work argues, reflecting on what is valuable to us, individually and collectively, is not a problem awaiting solution but part of what animates human life in the first place. The question of valuation is, at its core, a question of meaning, and meaning-making is central to the human condition.

This conviction brings this study close to particularist strands of ethical thought (for a comprehensive defence of such an account, see Dancy 2006), grounded in the view that the domain of ethics is too complex to be captured by universal principles. From foundational questions in metaphysics and epistemology, through questions about how these shape our understanding of value, to the application of ethical reasoning in concrete situations at different scales, ethical judgement is context-dependent, and no universal and fixed principle can therefore claim authority across these domains.

This complexity also means that any discussion of valuation needs to make its assumptions and commitments clear. Since this study both interprets existing approaches to morality and ethics and develops one of its own, this need for clarity applies both to how I understand other scholars’ terms and to how I use them in developing the virtue-ethical proposal of this study. As for others’ approaches to morality and ethics, wherever possible I read them from within

their own conception of the relevant term, and my critique should therefore be understood as internal to each approach unless stated otherwise.

As for the status of the characteristics of morality offered in this section, the purpose is not to advance a particular form of morality but to identify the target of the critique. In arguing for a virtue-ethical approach to theorising, the study sets itself against moral reasoning that precedes ethical thinking or proceeds in isolation from it. The definition of morality (1.4.1) therefore marks out the object of that critique rather than the framework the study seeks to defend, while the conception of ethics (1.4.2) belongs to the proposal itself.

The working definition of morality adopted here is accordingly a synthesis of several accounts of modern morality which, while they differ in some respects, converge on a set of core features. What they converge on is a principle-centred formation, and it is that formation, rather than moral philosophy as a whole, that I take as my object. Particularists reject the governing role of principles as firmly as I do (for an overview, see [Väyrynen 2011](#)), which is why they fall outside the definition rather than counting against it. The definition is thus methodological rather than substantive, as it sketches the phenomenon under examination without endorsing it, and without claiming that every moral philosopher is committed to it.

Unless otherwise indicated, the following working definitions apply wherever I discuss morality and ethics from my own perspective. Section 1.4.1 outlines what I take morality to encompass; section 1.4.2 turns to how ethics and ethical valuation are understood in this study.

1.4.1 The anatomy of the moral system

The conception of morality I sketch here is intended only to describe the position I am criticising, not a framework that I intend to defend.

In brief, morality subjects valuation to a formal method of reasoning aimed at producing determinate and binding judgements about right and wrong, and good and bad. More fully, morality is a system of principles for judging whether something is fundamentally right or wrong, good or bad. The term ‘morality’ can be understood either descriptively, as the code of conduct that a society, group, or individual actually endorses, or normatively, as a code that, under specified conditions, would be endorsed by all rational persons (Gert and Gert 2025). This study is concerned with the normative dimension.

Morality is not an individual phenomenon but a social one. It organises relations among entities that regard one another as capable of responding to moral demands and being held answerable to them. Moral agents therefore have a

framework of moral reasons that they use to assess one another's actions (as either moral or immoral). So, moral agents may disagree deeply on the substantive content of morality, as they often do, but what they share is the notion that morality matters.

From here, it further follows that one cannot properly be said to act morally or immorally towards entities so unlike us that they neither share that framework nor stand any realistic prospect of coming to share it. At this point, a critic might object by asking, for example, what of non-human sentient beings on Earth? Animal ethics has a well-established place in moral philosophy. Peter Singer ([1975] 2009), a utilitarian, is perhaps its best-known advocate, while Christine Korsgaard (2008), writing from a Kantian perspective, has rejected Kant's direct denial of equal dignity to animals and has argued, using Kantian reasoning, that animals should be regarded as ends in themselves. But my argument is that morality is not about the moral standing of such entities but about participation in a scheme of reciprocal accountability. This is not to say that it would be permissible to harm other beings or simply ignore them. It simply means that, because our relationship to them is non-reciprocal, it lacks the mutuality that obligations involve, and thus harm is still ruled out on other grounds, since it takes away another's capacity to value at all. Our relation to such beings is therefore more accurately described in ethical than in moral terms. The question of attributing a dignified status to others turns out, on this account, to be a question about ourselves – what characterises our being as human when we engage with entities with whom we share no moral code? Ethical valuation, as I conceptualise it, does not presuppose shared accountability. It is not about 'what we owe to each other' but about 'how we are towards each other', and it holds regardless of who or what the other is, since the question is not whether the other has standing to make a claim upon one. What ethical valuation requires is reflexivity, a disposition to hold oneself accountable for what one values. In other words, how one relates to others is inseparable from how one relates to oneself, since one's values also shape how one understands and evaluates one's own conduct.

Let us now consider the defining characteristics of morality.

Morality presents itself as non-contingent. This is a second-order claim rather than a first-order one. I am not claiming that morality is actually relative to whoever happens to be judging. Rather, I am describing how morality is understood by those who reason within it, and how the sources of their judgements – beliefs, preferences, and sensibilities shaped by different influences – can be hidden from them. This is not a metaphysical claim about whether moral knowledge exists independently of the human mind, a question that need not be settled here, nor is

it the epistemic claim that moral beliefs are systematically mistaken. This account follows Nietzsche's genealogy (1.3).

Morality is *solution-seeking*. Moral reasoning starts from the premise that there is an answer to every question. There may be several defensible answers, and choosing between them may involve some moral loss, but at least one moral answer is always available. This is tied to a desire for certainty, since moral reasoning seeks to settle a moral question and reach a point where nothing more needs to be said. Moral verdicts are therefore treated as final, and because they are final, they are considered in binary terms of right and wrong.

Section 1.2.1 located this presumption in Enlightenment rationalism and its extension of scientific method into the human and social spheres, and in the Kantian assumption that practical reason can supply a correct answer to every moral question.

The considerations morality invokes are further held to possess *intrinsic value*, carrying weight in virtue of what they are, of themselves, in themselves and for themselves. This does not exclude instrumental, pragmatic or functional reasons for maintaining a moral system, but such reasons are derivative. That moral judgement rests on the intrinsic worth of what is valued is a mainstay of moral philosophy (Zimmerman and Bradley 2025), and section 2.3.1 examines how Rawls's conception of justice depends on the intrinsic value of justice itself.

Together, intrinsic value and determinacy give rise to obligation. If we consider morality's requirements as carrying worth for their own sake, and if what it requires can be established, then the moral agent is bound to act accordingly. Importantly, this obligation is not an abstract idea; it concerns what we are actually required to do. Simply recognising that something is valuable does not, on its own, give us a reason to act. The principle that 'ought implies can' only limits what we can reasonably be expected to do; it does not create the obligation to act. Obligation is also the feature that critics have addressed most often. For instance, Anscombe (1958) notes that the grammar of obligation is embedded in morality, Nietzsche traces the sense of moral duty to an internalised relation of debt ([1887] 2007, §§II.4–6), and Williams argues that the obligatory nature of morality is internally contradictory (this will be taken up in more detail in section 3.2.1). Something recognisable as obligation nonetheless appears across the codes that societies actually endorse (Gert and Gert 2025), which is part of what makes it so difficult to dislodge.

Morality functions through *principles*, which are universal in scope. They are general formulations that apply across situations sharing a relevant feature, addressed to any agent to whom that feature applies, and from which more specific rules may be derived. Kant's ([1785] 2008) systematisation of morality

under a supreme law from which particular duties follow is a good example, because before the categorical imperative procedure can be applied to a case, that case must already have been described in morally relevant terms, and the framework itself does not supply that description (Herman 1993, 75). Morality, in other words, begins only once the question of relevance has been settled elsewhere.

Principles thus function as heuristics, condensing the results of past moral deliberation into general rules that can be applied efficiently to new cases. As heuristics, principles are useful. Deliberating directly from values to particular judgements is cognitively demanding, and our knowledge of any situation is inevitably incomplete. Thus, the critique developed in this study is not directed at principles as shortcuts, but at the governing role morality assigns to them. This is because, when principles become the starting point of judgement itself, they start shaping the meaning of the very values they are meant to express. This risks resulting in a self-justifying moral system in which objections are addressed by appealing to the governing principle, while values are interpreted in ways that preserve the system's consistency, ultimately producing epistemic closure, i.e., closing off the possibility of questioning the system from within.

To summarise, the search for solutions assumes that there is a right answer to be found; moral principles state what matters; intrinsic value makes that answer difficult to question; and obligation binds the agent to it. The moral system² therefore hides the fact that its values are contingent. My objection, then, is to deciding what is valuable before the agent experiences the particular situation. This is the common feature I find in moralism, realism, and agonism throughout the chapters that follow, and it is this feature that my proposal seeks to reverse in chapter 6.

1.4.2 On ethical valuation

In this study, ethics is understood as an *epistemic* enterprise concerned with how values acquire meaning through agents' interpretive and creative engagement with the world, and with human beings as the agents of that valuation. Where morality prescribes what one ought to do, ethics asks how anything comes to have value at all. Morality, in other words, is concerned with action, whereas ethics is

² It is worth recalling, however, that not every moral theory, nor every application of moral theory to politics, fits neatly within the characterisation given here. For instance, while Kant does, Rawls fits in a different way, since his principles are conceptual rather than substantive: he does not derive political principles from a prior moral doctrine, but the two principles of justice are moral principles nonetheless, universal in scope and generative of more specific requirements (see 2.2 for a detailed discussion).

concerned with meaning-making and with the being and becoming of the value-creator. While the broad question ethics asks is how we might live well, the question of how value is created needs to be asked before that. Ethics, in this sense, is therefore not a system for truth-seeking but a way of thinking about meaning, with this study focusing specifically on ethical valuation.

A careful reader will notice the terms related to thought throughout this study: moral reasoning, ethical reasoning, and ethical valuation, the last of these belonging to my own proposal for a creativity-centred virtue ethics. Let us therefore now see how they differ, on my view.

I consider reasoning in general to be an inferential process that involves drawing conclusions from premises and reaching judgements from different considerations. Good reasoning involves maintaining coherence and logic, holding different considerations in a clear relation to one another. Valuation, as the term is used in this study, is what makes something matter to us in the first place. It entails interpreting something, considering its relevance, and creating meaning, but it does not seek a judgement. Valuation is not inferential, in the sense of following a sequence of steps. Instead, it asks why something matters and what makes it so, relying on perception and attention, i.e., on noticing the qualities of something that make it meaningful to us. Valuation is tied to the person doing the valuing; the valuer's reasons are internal to them and entail an affective dimension. How we are disposed towards something – how we perceive and feel about it, and respond to it – is decisive for what that thing comes to mean for us. Valuation therefore draws on the valuer's internal motivations rather than on some universal standard.

Let us illustrate the distinction between reasoning and valuation using Iris Murdoch's well-known example of M&D in *The Sovereignty of Good* ([1970] 2014, 16–24)³. For Murdoch, attention is the core concept of her moral philosophy, denoting “truth-seeing and truth-discovering attitude and activity” (Caprioglio Panizza 2022, 156). The situation Murdoch describes is this. M feels hostility towards her daughter-in-law, D, whom she finds good-hearted but unpolished, “pert and familiar, insufficiently ceremonious, brusque, positively rude, always tiresomely juvenile” ([1970] 2014, 16–17). Over time, M's ‘vision’ alters (ibid., 17), and Murdoch describes the change as “the substitution of one set of normative epithets for another” (ibid., 18), brought about by “giving careful and just attention to an object which confronts her” (ibid., 17, original emphasis), so that she discovers D “to be not vulgar but refreshingly simple, not undignified but spontaneous, not noisy but gay, not tiresomely juvenile but delightfully youthful,

³ The interpretation of the example in terms of valuation is my own; Murdoch does not conceptualise it in these terms.

and so on” (ibid., 17). Nothing in the circumstances changes and, more importantly, neither does M’s outward behaviour, which was impeccable throughout.

M thus changed her disposition through attention⁴, which Murdoch describes as “the effort to counteract such states of illusion” (ibid., 16–17), and which consists in considering reality in normative terms (ibid.). Put otherwise, what M altered was her perception of reality by changing her disposition through the removal of her illusions. The attention that enabled this change was not directed towards the other, nor was it guided by an external signal or by any desire on M’s part to demonstrate a change in her behaviour. She engaged instead in an “internal struggle” (ibid., 22): “I may be prejudiced and narrow-minded ... Let me look again” (ibid., 17). The attention she gave was, first of all, directed towards herself.

There were, then, no externally motivated or instrumental reasons involved, nor was it formal reasoning that led M to change her perception of D. Nor can it be known exactly how the process unfolds in formal terms because, as Murdoch notes, “M’s activity is hard to characterize not because it is hazy but *precisely because it is moral*”⁵ (ibid., 22, original emphasis). The activity M engaged in was valuation. She had no conclusion in mind when the inner struggle began, and indeed no external motivator pushed her towards it. The endeavour was entirely her own and entirely private, so that there is no direct explanation of what moved her that an external observer would immediately recognise.

Such valuation, Murdoch notes, is an endless task, “something infinitely perfectible”, since “this new picture has built in the notion of a necessary fallibility” (ibid., 23). Her governing metaphor is vision, since what M changes is the way she sees, and this shift cannot be set out as a logical sequence because it is bound up with how she feels about her circumstances. On Murdoch’s example, then, valuation is neither the attribution of an arbitrary good-or-bad judgement to an object nor the creation of a fantasy to match a configuration settled in advance, but, in her words, a “progressive attempt to see a particular object clearly” (ibid., 23).

⁴ Attention is Murdoch’s central concept, both as a path to goodness and as a way of inquiring into moral truth. This study will draw on this concept in the context of virtue-ethical valuation, inspired by Murdoch, and will conceptualise attention as something that activates virtue and is present throughout the entire process of individual valuation, while also extending into the social and political sphere of self-creative democracy. This will be discussed in more detail in section 6.3.2.

⁵ Although Murdoch uses the term ‘moral’ throughout her work, she is an ethical thinker in the sense employed in this study, and linked with virtue ethics (see also [Blum 2023](#)), as her work is concerned with the character and being of persons, and she strongly rejects objectivist modern empiricism ([1970] 2014, 24–25).

Let us now construct a hypothetical case of reasoning on the basis of Murdoch's example, which is, of course, only one possible scenario. Reasoning ethically, M might ask what kind of person she needs to be to ensure a good family life; reasoning morally, she might ask what she owes her son's wife. Considering the reasons for D's behaviour, and drawing on her knowledge of the wider social world, she might come to see that the manners she expected of D belong more to her own social class than to social rules in general. She would then conclude, if she is reasoning morally, that D does not owe her those particular forms of behaviour, and that what she herself owes D is an invitation to take part in the family's life without having to put on a performance; and, if she is reasoning ethically, that kindness and benevolence must begin with herself. In either case, the conclusion is reached by moving step by step towards the judgement.

In valuation, M begins from epistemic doubt: what if my perception is incorrect and is the product of illusions created by the comfort of bias?⁶ The question she begins with is not one that yields a definitive conclusion, since circumstances change and M must attend continually, not because she wishes to do the right thing but because attention and the "loving gaze" (Murdoch [1970] 2014, 33) are part of her way of being. Valuation, in other words, begins in a suspicion directed at oneself, proceeds by attending rather than by inferring, and does not seek a definitive and final resolution.

In ethical reasoning, by contrast, the aim of bringing about an ethically good situation is approached through a rational line of thought. M might ask what kind of person she must be if there is to be a good family life, and she reasons her way, step by step, to an answer about which of her expectations are hers to revise. The aim of ethical reasoning is to reach a conclusion. Moral reasoning has the same aim, but it begins from a principle already in place – in this case, that the mutuality of a relationship requires everyone to be treated justly – and proceeds from this premise, and the justifications for such justness then issue in a judgement about what must be done.

So conceived, the central object of ethics is the human being rather than their action, since valuation occurs through an agent's experience and their interpretive engagement with the world, and human beings figure here as creators of value as much as evaluators of it. This notion situates the study within the continental approaches surveyed in 1.2.2, since what follows draws primarily on Nietzsche's account of valuation.

⁶ To emphasise, this is one example only. Murdoch is a moral realist, and her perspective reflects a seeking of the good in reality, so that a valuer working from a different paradigm might begin from reasons other than the uncovering of illusion.

To elaborate on this, let us recall the discussion of agency and status from the previous section. The working conception of morality treats answerability as essential to a moral relationship, which means that it cannot apply the language of obligation to beings who are sentient but so different from us that they cannot hold us accountable. Ethical reasoning, by contrast, does not depend on reciprocity, so it allows us to regard others as mattering, regardless of whether they share our moral vocabulary. What it does require, as I set out in 1.4.1, is reflexivity, understood as a disposition through which one values something for which one is oneself answerable. This follows in turn from ethics being concerned with the being and becoming of what I term the ‘value-creator’, or ‘esteemer’, to borrow Nietzsche’s ([1883] 2006, *On a Thousand and One Goals*) term. So understood, to think about others in ethical terms is therefore to think about how we are, and how we might be, towards them, rather than about what we owe them or what we stand to gain from the relationship.

On this view, the meaning-making that shapes being and becoming is reflected in dispositions, which is why ethics issues no code of conduct, since although it can think in general and even universal terms, it does not seek to establish universal guidance for action. Its concern is with attention, perception, and responsiveness rather than with particular claims or justifications, and it is primarily an understanding of how that meaning is formed and what it yields for how we live together, rather than a matter of coherence or logical consistency. Because persons stand at the centre of the inquiry, any adequate account of ethics would have to account for human character traits, which is why the question of how the virtues operate in the meaning-making of values, and what dispositionality involves, is taken up in chapters 5 and 6.

So ethical valuation does not regard the cognition of value as a matter of logic alone. It takes it instead to be a matter of active interpretation – of the circumstances, of oneself, and of the object of valuation – in which affective motivation is equally relevant. Such motivation answers neither to instrumental interest nor to the pull of a preset standard, and it is guided by the view that affect is part of human cognition (Helm 2007; Slote 2010).

Ethical valuation therefore involves two cognitive dimensions, which I term sense and sensibility. Sense concerns the more rational and interpretive process of sense-making, i.e., what values mean and their intelligibility, while sensibility concerns how their significance is felt. If these were two actors, the Sense-Maker would ask, ‘What does this mean?’, while the Sensibilist would ask, ‘What is its significance?’

Sense and sensibility also include the dimension of affective cognition. Psychological research suggests that while instrumental motivation moves a

person towards a particular goal, affective cognition is what prompts us to move in the first place, generating a desire to act or to change before any reason for doing so has been articulated (Silvia 2008). Helm (2007) makes the same case on conceptual grounds, arguing that evaluative judgement requires active emotional engagement if the divide between cognition and conation is to be bridged. We may not know why we want something, and yet emotion underlies other forms of cognition, moving us to reason about the thing that has taken hold of us, and even to ask whether our interest in it has any rational warrant. Rational reasoning is, in this sense, itself set in motion by an intuitive and affective dimension.

I understand ethical valuation, then, as concerned first of all with relevance, that is, with whether, why, and how something matters within the social situation that calls for valuation, and indeed with whether that situation calls for valuation or carries an ethical dimension at all. Relevance of this kind is worked out through meaning because it is meaning that makes things matter to us. What has worth for us is not valuable in virtue of some prior knowledge of its worth; it comes to be valuable through the meaning it acquires for us. Human beings are, on this account, not creatures of logic alone but of experience and feeling, as the diverse and often messy texture of our social relations shows, nowhere more so than in matters of valuation. But I take as a starting commitment of ethical thought that such messiness is not an error to be resolved, whether by moral codes or otherwise. The messiness of social life does not indicate confusion but a lasting complexity, which is precisely what ultimately makes social relations imaginatively productive (see chapter 6 for more).

A critic might ask at this point in what sense ethics, so conceived, is normative. If it yields no code of conduct, formulates no principles, and reaches no verdicts, what kind of normativity does it entail? The answer is that to hold that values are created and re-created by those who live by them is already to make a claim about how we ought to engage with the values by which we live. If we are not merely given values but participate in creating and sustaining them, then how we attend to, interpret, and respond to what matters is itself open to normative assessment. Ethics, in the form outlined here, is therefore normatively interpretivist and not merely descriptivist, since it is concerned not only with how meaning is formed but also with how we ought to take part in forming it.

To engage in ethical valuation therefore entails shaping what one attends to and what one notices, and consequently what one is able to see, akin to what Iris Murdoch ([1970] 2014) discusses as the ethical importance of attention. Attention and valuation concern how one lives, rather than simply how one follows rules for living. The normativity at issue here is therefore not primarily prescriptive; it does not begin by asking what one ought to do, but how one ought to perceive,

interpret, and respond to what matters. This will be discussed further in chapter 6. Accordingly, I distinguish between being normatively engaged through perception and response and being prescriptive about what one ought to do.

2. POLITICAL MORALISM

2.1 WHAT IS POLITICAL MORALISM?

Political moralism is a term coined by Bernard Williams (2005) to distinguish between a mode of theorising aimed at establishing a good social order and a perspective in which politics possesses its own paradigmatic framework, generating distinct aims and concepts. According to this view, moralism denotes an approach to political theory in which politics serves as a means to realise certain moral goods, such as justice.

Political moralism is most closely associated with John Rawls (1996, [1971] 1999). Although the view also encompasses other liberal theorists, such as Dworkin (2011) and Scanlon (2000), and may extend to adjacent normative approaches such as communitarianism insofar as they treat moral values as prior to politics (Rossi and Sleat 2014), Rawls remains its most influential representative. He is therefore also the primary target of critiques from political realism (Leader Maynard 2025, 13) and agonism (Mouffe 2005), discussed in the following chapters. The discussion below focuses on Rawls to avoid conflating different ‘dialects’ of political moralism.

This chapter examines the moral dimension of Rawls’s framework, focusing on how moral considerations inform its political arrangements and on the moral commitments that underlie them. My aim is not to provide a comprehensive account of the theory, but to illuminate the relationship between its central conceptions and the moral foundation of political liberalism, that is, to explore the implications of theorising about political arrangements from a moral starting point. For Rawls, morality is what makes politics, as opposed to mere coercion or manipulation, possible. He holds that we must have reasons for engaging in public dialogue that exclude some of our most deeply held commitments as justificatory arguments, particularly under conditions of enduring moral and cultural pluralism, and he offers a framework for understanding how the very reasons that lead us to bracket certain private commitments in public discourse are also the ones that give rise to those commitments in the first place. In his own words, political liberalism is about “conceptions of a free and equal citizen, of the legitimacy of the exercise of political power, of a reasonable overlapping consensus, of public reason with its duty of civility, and of stability for the right reasons” (1996, xli). As we shall see, the “right reasons” are those grounded in considerations of justice (1996, xli, 391), and the conviction that social stability is truly stable only when it rests on them is, I take it, the core commitment of Rawlsian moralism.

Before entering Rawls's world, we must first clarify the idea of moralism itself: what does it mean to describe a political theory as moralist? Jonathan Leader Maynard provides a helpful definition:

“the view that proper practical normative judgements about politics – about what political actions or norms or institutions are good, right, or permissible – depend, in some sense, on moral judgements... Political moralists consequently hold that moral reasoning is a necessary, though not sufficient, task of political theory – whether as a positive basis for justifying our political judgements or for the more critical interrogation of the moral assumptions they involve... However, this does not imply that the only reasons for political action are moral reasons” (2025, 14)

This definition is formulated in contrast to political realism, which is discussed in chapter 3. Realism considers politics to have a distinctive kind of normativity of its own. On this view, concepts such as legitimacy, order, and stability have value in themselves, rather than being derived from prior moral or other normative commitments. The central divide between moralism and realism therefore concerns the order of justification, since moralism gives moral reasons epistemic priority, whereas realism denies that they hold such a privileged position.

But what distinguishes a moral judgement from any other kind of normative judgement? Leader Maynard argues that moral judgements “represent a class of normative judgements distinguished by the kinds of *reasons* involved” (2025, 15, emphasis added), and are therefore distinct from, for example, instrumental or aesthetic judgements. He further suggests that the most concise and encompassing way to understand morality is through the notion of “other-regardingness” (ibid.), whereby moral considerations guide us to “consider what should be done with regard to the interests and status of others” (ibid.). As we shall see in the concluding section, however, this criterion is less neutral than it initially appears, since defining morality in terms of its orientation towards others already presupposes a particular moral outlook.

A clarification about the term ‘moralism’ is in order before we proceed. First, I use ‘moralism’ as a descriptive term, not as a criticism. The term was coined by realists and carries a critical meaning in that literature, but the discussion that follows is as much a defence of Rawls against the realist critique as it is an argument against him. Second, I use the term more broadly than the realists do. For Williams, a theory is moralist when it gives moral principles priority over political judgement. In this study, however, the more basic question is whether a theory decides in advance what is valuable.

Moralism in the realist sense refers to the view that political arrangements should be evaluated, at least in part, by moral standards. As we shall see, this can take either a substantive or a conceptual form. Substantive moralism (which, as I will argue below, Rawls does not endorse) treats politics as a vehicle for realising a substantive moral doctrine that determines what is valuable in the particular political order, such as a certain conception of justice. Rawls, by contrast, is a conceptual moralist, as he settles in advance the conditions under which citizens determine the content of justice, namely the two moral powers, the original position, and the veil of ignorance. Neglecting this distinction between substantive and conceptual moralism, I argue, is why the realist critique does not fully capture what is at stake in Rawls (see chapter 3).

Having introduced political moralism in general, we now turn to Rawls's political liberalism as its most prominent exemplar: we begin by discussing the particular ways in which Rawls relies on morality in his theorising (2.2) before turning to the machinery of the theory itself: his conception of justice (2.3), his account of politics and liberal legitimacy (2.4), and his assumptions about political actors and the moral and epistemic powers he ascribes to them (2.5). The chapter then considers the objections his theory has attracted concerning feasibility and power (2.6) and concludes by reflecting on what can be learned from Rawls's moralism (2.7), where I argue that theorising about political arrangements loses much of its significance unless it is guided by a value-based aim that gives individuals and societies reason to strive for a harmonious (though not harmonised) form of collective life.

2.2 RAWLS'S CONCEPTUAL MORALISM

2.2.1 Rawls on morality

In his essay *The Independence of Moral Theory* (1974), Rawls distinguishes between moral philosophy and moral theory. The former denotes the broader field, including questions about the nature of morality and moral thought (including, for example, metaethics), whereas moral theory concerns the structure of moral content. In Rawls's words, moral theory is "the study of substantive moral conceptions, that is, the study of how the basic notions of the right, the good, and moral worth may be arranged to form different moral structures" (ibid., 5). While Rawlsian moral theory does not ask which moral conception – for example, Kantian, utilitarian, intuitionist, or perfectionist – is true, its object is nonetheless substantive, since it examines competing approaches that seek to establish principles for judging what is right and good without adjudicating among those conceptions. It maps the structures that moral

conceptions take and considers them in relation to human psychology, or how they fit “our moral sensibilities and natural attitudes” (ibid.). Crucially, Rawls treats the question of moral truth as something to suspend rather than settle, arguing that “since the history of moral philosophy shows that the notion of moral truth is problematical, we can suspend consideration of it until we have a deeper understanding of moral conceptions” (ibid., 7). Put simply, Rawlsian moral theory examines the structures of moral reasoning that frame the search for moral truth, but it does not ask whether such truth is absolute or can be known at all.

This is also where Rawls locates what this study terms ethical reasoning (see 1.4), though the fit is somewhat loose. Rawls occasionally uses “methods of ethics” (1974, 14) as a synonym for intuitionism and utilitarianism, referring to them as “moral conceptions” (ibid.), rather than as a term marking a distinct technical category. Accordingly, we can assume that, by ethics, Rawls means the broader field of moral philosophy.

For Rawls, moral theory is comparative and structural, since it seeks reflective equilibrium between principles and considered judgements rather than the discovery of a first-order moral truth beneath them. Formal conditions such as generality, universality, and publicity are, he insists, “best viewed as simply very general properties” (1974, 10–11). The task of moral theory is therefore to investigate “not just a question of meaning but of how the whole theory that results fits together”, comparing alternative moral structures and examining both their logical properties and their overall coherence (ibid.). Rawlsian moral theory, in short, is more akin to architecture than archaeology, as it constructs and compares moral frameworks rather than excavating a truth presumed to lie beneath them.

Rawls also emphasises that moral theory is independent of other branches of philosophy, such as epistemology, the philosophy of language, and the philosophy of mind. A moral theorist need not resolve broader philosophical questions, for example, whether morality is metaphysically objective, before engaging in moral theorising, because the study of substantive moral conceptions “has its own distinctive problems and subject matter that requires investigation for its own sake” (1974, 6). In other words, moral theory’s independence is not instrumental, derived from its usefulness elsewhere, but is instead a claim about its disciplinary autonomy. From this, it follows that the usefulness of a moral theory must also be judged from within the theory itself, based on whether it is internally coherent rather than whether it can be validated from outside. A Kantian theory, for example, cannot be meaningfully criticised from a utilitarian perspective. Starting from the idea of Kantian autonomy, a moral theory can instead address real problems that arise in social life, including the stability of a

well-ordered society (ibid., 13), because it would be unhelpful to wait until meta-questions such as whether morality can be true or false have been settled before developing a substantive moral theory, especially since Rawls avoids metaethical discussions on the source of moral knowledge.

In *Political Liberalism*, Rawls clarifies the distinction between a concept and a conception of justice:

“Roughly, the concept is the meaning of a term, while a particular conception includes as well the principles required to apply it. To illustrate: the concept of justice, applied to an institution, means, say, that the institution makes no arbitrary distinctions between persons in assigning basic rights and duties, and that its rules establish a proper balance between competing claims. Whereas a conception includes, besides this, principles and criteria for deciding which distinctions are arbitrary and when a balance between competing claims is proper. People can agree on the meaning of the concept of justice and still be at odds, since they affirm different principles and standards for deciding those matters. To develop a concept of justice into a conception of it is to elaborate these requisite principles and standards.” (1996, 14, fn. 15)

Accordingly, Rawlsian moral theory does not ask ‘What is just?’ but examines the structure of justice, or the concept of justice, whereas the question of what counts as just belongs to a particular conception of justice. Rawls’s discussion of moral theory therefore focuses on the conceptual framework within which judgements about justice are made, rather than on the specific content of those judgements.

Rawls inherits much from Kant, but not everything. As Southwood (2009) observes, Rawls is a social contractarian concerned with political agreement rather than a deontologist seeking universally binding moral principles. In *A Kantian Conception of Equality* ([1975] 2001), Rawls specifies what he takes from Kant and what he does not. Rather than retaining Kant’s dualisms (necessary/contingent, form/content, reason/desire, noumena/phenomena) in the sense Kant gave them, he reformulates them “within the scope of an empirical theory” (ibid.). He illustrates this through Kant’s idea of negative freedom and his own device of the original position, since on this view we are free when our choices are not determined by external forces or accidental circumstances. Applied to politics, this means that principles of justice should not be chosen because they happen to benefit particular individuals or reflect contingent historical, social, or natural circumstances. Rather, they should be chosen independently of such contingencies (ibid., 265).

Later, in *Political Liberalism*, Rawls distances himself even further from Kantian moral constructivism (1996, 99–101). As he writes, “we must, of course, reject

Kant's constitutive autonomy" (1996, 100), that is, the claim that practical reason itself constitutes an independent order of moral value, since "transcendental idealism and other such metaphysical doctrines play no role" in organising the political conception (ibid.). Rawls thus rejects the Kantian view that moral normativity governs every sphere of life, but what he keeps is Kant's conception of the moral person as rational and capable of forming, revising, and pursuing moral ends (ibid., 99–100; see also 2.5.1).

What Rawls inherits from Kant, then, is the conception of the person, but not the conception of moral duty. The moral powers Rawls attributes to citizens are capacities rather than fixed commitments (ibid., 103–105), and the duty of civility that follows from them (see 2.4.2) is political rather than comprehensive. Because the duty of civility relies on rational deliberation, it keeps moral judgement apart from anything that might get in the way, such as emotion, social norms, or personal relationships. As a result, the system is hard to criticise from the inside, and its underlying mechanisms can only be identified from an outside perspective. Thus, political principles should not be grounded in fixed rules imposed from outside politics; they should instead be shaped through public debate and the search for common ground.

Finally, and to re-emphasise, Rawls is not engaged in moral theorising, or indeed in moral philosophy, in *Political Liberalism* itself, since in that work he neither compares substantive moral conceptions nor seeks reflective equilibrium among them. He merely uses and clarifies the concepts and distinctions already developed by moral theory as resources for further political construction, without repeating the comparative work of moral theory itself.

Having outlined Rawls's understanding of morality, and in particular what I call his *conceptual moralism*, we can now turn to the role this conception plays in his project of political liberalism.

2.2.2 Rawls's conceptual moralism in political liberalism

Rawls's project, spanning *A Theory of Justice* ([1971] 1999) and *Political Liberalism* (1996), seeks to reconcile two competing traditions in democratic thought, one deriving from Locke and prioritising the liberties of the moderns, the other deriving from Rousseau and emphasising the equal political liberties of the ancients (1996, 4–5).

The solution Rawls proposes to resolve the tension between liberty and equality is a contractarian one (Cudd and Eftekhari 2025). Through a process of "reflective equilibrium" (Rawls [1971] 1999, 42–44), parties seek agreement on the principles of justice that should govern the basic structure of society while situated in the "original position", where they lack knowledge of their own social position,

talents, or comprehensive commitments. This hypothetical approach aims to ensure fairness by requiring people to choose the principles behind a “veil of ignorance” (ibid., 118–123). This prevents their decisions from being influenced by factors they happen to be born into, such as their social position, wealth, or health. The aim of this procedure is to secure equal respect for persons as free and equal citizens and to establish a shared political framework within which collective decisions about the terms of social cooperation can be made.

The parties are assumed to adopt two principles of justice. The first holds that “each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all”, while the second requires that “social and economic inequalities are to be arranged so that they are both: (a) to the greatest benefit of the least advantaged, consistent with the just savings principle, and (b) attached to offices and positions open to all under conditions of fair equality of opportunity” ([1971] 1999, 266).

Methodologically, the original position is a thought experiment rather than an empirical or prescriptive exercise. As Rawls emphasises, it is “not, of course, thought of as an actual historical state of affairs” but rather “a purely hypothetical situation characterized so as to lead to a certain conception of justice” ([1971] 1999, 11). Its purpose is to identify the principles that reasonable parties would choose under fair conditions, not to derive a conclusion that follows deductively from its premises.

Political Liberalism (1996) is an extension of *A Theory of Justice*, adding a question left open by the earlier work: how can a just and stable society remain stable when citizens hold different and incompatible, yet reasonable, comprehensive doctrines? Rawls calls this condition “reasonable pluralism” (ibid., 36–37). In *A Theory of Justice*, Rawls had based stability on the idea that citizens’ conception of justice would align with their personal conception of the good, but later he came to see this assumption as unrealistic (ibid., xviii). However, the core claim of justice as fairness remained the same, namely that it is concerned with the basic structure of society, meaning the way its main political, social, and economic institutions are organised (ibid., 257–258).

So, Rawls’s political liberalism retains the core of justice as fairness, while reframing the question it is meant to answer. Rawls rejects the assumption of a shared comprehensive doctrine, on the basis that persistent disagreement among reasonable but incompatible doctrines is the permanent result of free institutions rather than a defect to be overcome (1996, xvi–xvii, 36–37). Accordingly, the question is no longer only what justice requires, but how a just society might remain stable in the context of reasonable pluralism, giving rise to three revisions. First, the conception of justice is presented as political rather than

metaphysical, in that its scope is limited to the basic structure and it is freestanding, that is, not derived from any comprehensive doctrine (ibid., 11–15). Second, its acceptance is to be secured through an overlapping consensus, in which citizens affirm the same political conception from within their several doctrines and for their own reasons (ibid., 133–172). Third, the stability thereby achieved counts as genuine only when it is stability for the “right reasons” (ibid., xli, 391) – the reasons of justice, rather than a fortunate balance of forces. This, then, is the model Rawls offers for organising the political, social, and economic institutions of a pluralistic society.

Being political in this sense does not, however, place such a conception of justice outside the moral sphere. Rawls insists that both justice and its subjects are not merely political but also moral in nature: “The citizen is, of course, a moral agent, since a political conception of justice is, as we have seen, a moral conception” (1996, xlv). What he does not fully clarify is what constitutes the moral dimension of this political conception – that is, how moral considerations inform political liberalism. Nor does *A Theory of Justice*, for all its comprehensive account of justice, thoroughly explore the moral foundations of its claims. This is what the present chapter aims to trace, starting with the point where the construction itself begins, i.e., the parties to the original position and how they are assumed to be.

The original position models citizens as free and equal moral persons, and it treats them as having two moral powers. The first is the capacity for a conception of the good, “the capacity to form, to revise, and rationally to pursue a conception of one’s rational advantage, or good”, where the good refers to what is “valuable in human life”. The second is the capacity for a sense of justice, “the capacity to understand, to apply, and to act from the public conception of justice which characterizes the fair terms of social cooperation” (Rawls 1985, 233). A citizen’s reasonableness therefore already presupposes a moral sensibility, since citizens are treated as moral agents precisely by virtue of these powers (see 2.5.1).

These moral powers ground a distinctively political obligation, namely the duty to seek consensus on acceptable social arrangements, so that where there is good reason to accept an agreement, fairness requires that it be accepted (Rawls 1996, 16). The duty runs both ways, binding also the party who makes the offer, since “in offering fair terms we must reasonably think that citizens offered them might also reasonably accept them. And they must be able to do this as free and equal, and not as dominated or manipulated, or under the pressure of an inferior political or social position” (ibid., xlv).

What emerges is a contract binding morally empowered citizens together through the principle of fairness in pursuit of just arrangements. This is a “social

contract” (Rawls 1996, 258), one that Southwood (2009) distinguishes from the stricter *moral contractualism*, whose most influential formulation is Scanlon’s (2000). Ashford and Mulgan (2018) identify two key features of this distinction. First, Rawls’s social contract is more deeply Kantian, seeking principles that everyone would agree to rather than principles that no one could reasonably reject (for Rawls’s own comment on Scanlon, see 1996, 49 fn. 2). Second, Rawls’s social contract is clearly political, concerning the socio-political arrangements to be agreed upon rather than the derivation of moral norms as such. On Ashford and Mulgan’s reading, these features render the Rawlsian offer of agreement merely instrumentally motivated, as a means of persuading others to accept arrangements that serve one’s own interests, whereas the moral contractualist’s (e.g., Scanlonian) citizens are moved by an intrinsic desire to justify themselves to others.

To briefly comment on Ashford and Mulgan’s inference, I accept the first point, which indeed makes Rawls the less moralistic of the two, since acceptance sets a higher demand for morally neutral justification than reasonable rejection does. But the inference drawn from the second rests, I argue, on a subtle misreading of Rawls. On a closer reading, we see that Rawlsian citizens are moved by their shared “sense of justice” (1996, 103–104), which arises from their moral powers (see 2.5.1). In my reading, this means a shared belief that society should be just, which both supports and is reinforced by fair procedures, rather than expressing a particular view of what justice should require. Without this sense of justice, any stability those procedures generate would not be stability for the “right reasons” (ibid., xli, 391), and so would not be genuine stability at all. Thus, Rawls does not require that citizens collectively establish the good on which they are to agree; he asks only that they “reach free and informed agreement on questions of constitutional essentials and basic matters of justice” (ibid., 128). That he does not specify what counts as basic is precisely the point, since the question is left for citizens themselves to settle.

What this means is that the obligation generated by Rawlsian consensus is political, not moral in the sense of deriving from a comprehensive doctrine. As Rawls writes, “The worth of citizens in a well-ordered society is always equal because in such a society everyone is assumed to comply with just institutions and to fulfill their duties and obligations moved, when appropriate, by a sufficiently strong sense of justice” (1996, 280 fn. 16). Morality matters here only insofar as it motivates citizens to act in accordance with the political conception of justice. The sense of justice is therefore not a commitment to substantive moral truth, but a shared recognition of the value of justice and of political cooperation as the means of creating and sustaining a just society.

Rawls's moralism, then, is conceptual rather than substantive, in that it seeks consensus on basic matters of justice while assuming that citizens share the belief that justice is the foundation of society, since they are first and foremost moral persons who enter into cooperation by virtue of their moral commitments. In Rawls's own words:

“the object of consensus, the political conception of justice, is itself a moral conception. And second, it is affirmed on moral grounds, that is, it includes conceptions of society and of citizens as persons, as well as principles of justice, and an account of the political virtues through which those principles are embodied in human character and expressed in public life. An overlapping consensus, therefore, is not merely a consensus on accepting certain authorities, or on complying with certain institutional arrangements, founded on a convergence of self- or group interests. All those who affirm the political conception start from within their own comprehensive view and draw on the religious, philosophical, and moral grounds it provides” (1996, 147).

The distinction between conceptual and substantive moralism is crucial to this discussion. As the next chapter also argues, it underpins what I take to be the realist misreading of Rawls and explains why the realist critique ultimately misfires. It is therefore worth briefly recalling the distinction discussed in 2.2.1. So, the concept of justice is not the same as its content or, in Rawlsian terms, its conception. As a concept, justice is empty of substantive content and therefore cannot itself be evaluated as just or unjust. Rawlsian justice is moral only insofar as citizens are conceived as possessing moral powers, as free and equal persons, and as ends in themselves. Rawls arrives at the political conception through moral motivation rather than moral truth. Therefore, morality functions as a conceptual framework, which supplies ideas such as fair terms of cooperation, reciprocity, and moral personality. In this sense, conceptual moralism structures the space within which morally evaluative questions can be asked, rather than determining their substantive answers.

The consensus, in other words, is conceptually, but not substantively, moral at both ends. Its object is a moral conception of justice, with its political conception being derived from moral motivation rather than from procedural rules alone. And its grounds are moral, since each citizen affirms justice from within their own comprehensive view, drawing on the moral resources that view provides: “each view supports the political conception for its own sake, or on its own merits” (1996, 148), so that a Lutheran affirms justice as fairness for Lutheran reasons, a communitarian for communitarian ones, and so forth. The agreement is therefore not a merely strategic *modus vivendi*. This also answers critics who charge that Rawls is not moral enough, his foundations too weak to sustain

anything beyond *modus vivendi* stability (Habermas 1995; Cohen 2008). For Rawls, a conception of justice holds when citizens sharing a democratic political culture endorse it. However, such a culture is itself morally structured, since it affirms a plurality of comprehensive doctrines on the understanding that each has a moral core, and it is this shared affirmation of morality's relevance, rather than any convergence of interests, that makes cooperation, and thus politics, possible in the first place.

As with any methodological commitment, Rawls's theory carries presuppositions of its own. First, the fairness of the original position assumes participants who wish above all to avoid being left worst off; that is, Rawls does not consider citizens who might accept risk for greater gain at others' expense. The theory thus presumes a citizenry that has, pre-politically, rejected the possibility of such risk-taking. Second, and more generally, the Rawlsian agreement relies on the moral and epistemic powers (2.5.1–2.5.2), which are thoroughly cognitive in character. Rawls assumes these as pre-political dispositions, with the meanings and reasons of political justice emerging from citizens' comprehensive affirmations, but how the dispositions themselves are formed is left largely open, beyond the suggestion that just institutions cultivate them in those who grow up under them (1996, 142). And since the principles are set in advance, they limit which dispositions are admitted to public life at all, so that, as Rawls himself notes, the *unreasonable* are excluded (1996, 64 fn. 19). This presumes a very particular kind of citizen – namely, a liberal one – and leaves unanswered the question of how a non-liberal society might become liberal, if the principles apply only where broadly liberal dispositions already obtain.

2.3 JUSTICE

2.3.1 Justice and its intrinsic value

The first question I wish to address concerns the axiological status of justice: is justice, for Rawls, an end in itself, or a means to some other end? He writes: “while a political conception of justice is, of course, a moral conception, it is a moral conception worked out for a specific kind of subject, namely, for political, social, and economic institutions” (1985, 224). The central issue, then, is whether justice is valuable *instrumentally*, i.e., because it is necessary for political systems to function properly, or whether justice is valuable in itself, with political structures serving as a means of achieving it. Also, does Rawls conceive of justice as shaping the design of institutions, or primarily as a normative standard by which political arrangements are to be assessed? The answer appears to be that justice is intrinsically valuable insofar as it is not merely a political conception,

but also a moral one (ibid.). As Rawls emphasises, it is “the principles of justice that specify the fair terms of cooperation” (1996, 77), and not the reverse. Moreover, Rawls explicitly links justice to intrinsic value when he states: “such a conception of justice, in order to be a moral conception, must contain its own intrinsic normative and moral ideal” (1996, xlv, emphasis added). Likewise, philosophers often see intrinsic value as important to moral judgement (Zimmerman and Bradley 2025), insofar as something is considered morally good in itself rather than because of some external benefit.

Theories of value distinguish between two senses of intrinsic value, since something can be valuable in itself or valuable for its own sake (Rønnow-Rasmussen 2015). This distinction can also be understood in terms of how value exists vs. how we come to understand it, i.e., as ontological and epistemological aspects of value. So, when something is valuable for its own sake, it is about how we understand what a particular value means in practice. For example, happiness for its own sake requires us to experience happiness. Until we encounter and reflect on happiness, it remains an abstract idea. Lived experience, therefore, is one of the most important ways of understanding what a value means and what it involves.

Yet an epistemic question remains: how do we know that something is intrinsically valuable? Even if we intuitively recognise “self-expression, freedom, peace, [and] security” (Frankena 1973, 87–88) as intrinsically valuable, on what basis do we regard these goods as intrinsically worthwhile, or at least as valuable for their own sake rather than merely as means to something else? Frankena offers the example of truth, suggesting that “truth is not itself intrinsically good. It may not even be known” (ibid., 89), and what is good in itself “is knowledge of or belief in the truth... what is intrinsically good is the contemplation or experiencing of them” (ibid., emphasis added).

In other words, purely structural intrinsicism is meaningless in isolation, because our intuitions about what is valuable for its own sake emerge through experience rather than from objective knowledge or purely abstract rationality. This implies that moral judgements – and, by extension, moral norms such as those concerning justice – cannot be entirely abstracted from moral agents or considered independently of them. Our ‘sense’ of what is moral has an experiential dimension that cannot be reduced to pure rationality, understood as decision-making based solely on factual evidence. Rather, experience involves individuals perceiving and interpreting observations, thereby turning abstract phenomena into meaningful understanding.

This point will become important later in the chapter, when we examine Rawls’s own conception of the sense of justice, which he understands primarily in

cognitive terms (2.4). But more broadly, this account of intrinsic value concerns the sources of human motivation. The value at issue is therefore intrinsic to the motivational structure that Rawls attributes to citizens (1996, 124), rather than to a realm of values that could be discovered independently of human agents (see section 2.4.2 on moral truth as unowned).

When Rawls asserts that justice is a political, yet also a moral, concept (1985, 224), by virtue of its possessing an intrinsic moral ideal (1996, xliv), he is articulating the structural dimension of justice's intrinsic value. But does Rawls ascribe *intrinsic meaning* to justice? In other words, how is Rawlsian justice valuable for itself? Why, ultimately, should we care about justice?

To address this, we must consider Rawls's Kantian conception of the person, where individuals are treated as ends in themselves ([1971] 1999, 289). The intrinsic value of justice stems from its concern with persons as free and morally equal. As a political – but not comprehensive – moral ideal, justice is intrinsically valuable because it recognises citizens both as subjects and as objects of justice. To affirm their moral equality is to acknowledge their intrinsic worth, and with this, the meaning of justice is realised.

Moreover, justice is not concerned with individuals in isolation, but with *relations* between persons. For Rawls, this is captured by the criterion of reciprocity, according to which society is a fair system of cooperation when citizens offer terms they reasonably believe others, as free and equal persons, could reasonably accept (1996, xliv). Such reciprocity presupposes respect for others as moral equals.

The significance of something being intrinsically valuable – and thus requiring no further justification – lies in its capacity to provide a source of meaning and a reason to pursue such ends. In this light, the intrinsic value of justice defines the very *purpose* of politics, which is not only to secure social stability without resorting to violence or coercion but to do so on terms that all affected parties could reasonably accept, with no group treated as morally superior to another. Politics is therefore essential even in situations where outcomes might be achieved more efficiently through coercive means, because our primary moral obligation is to respect each individual's intrinsic worth and dignity even if this entails a loss of utilitarian efficiency or other instrumental gains.

Moreover, the political conception of justice has a considerably narrower scope than moral or religious conceptions of justice. This is because morality and religion are components of comprehensive doctrines, which:

“include conceptions of what is of value in human life, ideals of personal character, as well as ideals of friendship and of familial and

associational relationships, and much else that is to inform our conduct, and ultimately our life as a whole. A conception is fully comprehensive if it covers all recognised values and virtues within one rather precisely articulated system; whereas a conception is only partially comprehensive when it comprises a number of, but by no means all, non-political values and virtues, and is rather loosely articulated” (1996, 13).

In this respect, a political conception of justice does not extend to all dimensions of comprehensive doctrines, so it does not prescribe what individuals ought to value beyond the bounds of public reason and political association.

2.3.2 Why justice?

Thus far, we have discussed Rawls’s political conception of justice as rooted in the intrinsically valuable moral conception of justice. Let us now examine more closely what this intrinsic worth entails in order to understand how Rawls approaches the content of justice.

The Rawlsian concept of justice as fairness encompasses the idea that the political framework ought to ensure fairness in basic liberties, opportunities, and the distribution of resources, with inequalities structured so as to benefit the least advantaged members of society (1996, 271).

Rawls considers these requirements to characterise just basic institutions, as he expects that these are the principles that “free and equal moral persons, in a situation that is fair between them, would adopt for the purpose of regulating that structure” (1996, 271).

The realist critique of Rawls stems from the argument that he is not merely concerned with justice as a general concept, but rather with a particular conception of justice that political institutions are expected to implement (Raekstad 2021; Thomas 2017). Central to this critique is the difference principle, which is often seen as privileging the least advantaged as a particular category of citizens over others. But should this really be viewed as a form of moral contamination in politics? Realists maintain that conceptions of justice do not emerge *ex nihilo*; instead, they are historically situated and shaped by contingent social, cultural, and political meanings (Rossi and Sleat 2014). From this perspective, Rawls is not offering a neutral or universal account of justice, but rather advocating a particular normative framework for determining what ought to be considered just.

However, does the realist critique hit its mark? I argue that it does not. This is because Rawls does not impose a fixed, substantive account of justice (e.g.,

equality, liberty, solidarity) as the moral foundation of politics. Rather, he articulates the meaning of justice's intrinsic value, i.e., *why* justice is morally foundational to politics, not what its specific content must be. In contrast to an instrumentalist view – such as a utilitarian justification, which values justice insofar as it promotes certain outcomes – Rawls grounds justice in its moral significance as a condition for political legitimacy among free and equal citizens. Importantly, Rawls does not prescribe exact standards or criteria for what counts as just; instead, the content of justice is to be determined through a political process aimed at achieving an overlapping consensus. As Thomas observes: “Rawls does not think of an overlapping consensus as a set of pre-given answers, but rather as a way of framing our problems” (2017, 14).

Justice, in Rawls's account, is not conceived as a specific model of social life, but rather as the means through which political stability is achieved – crucially, stability achieved for the *right reasons* (1996, 391). That is to say, stability is not valuable in itself, nor should it be pursued merely for the sake of order; it must be underpinned by moral justification. For this reason, the *right reasons* that ground justice cannot be perfectionist ideals or abstract moral absolutes. Instead, they are shaped by the capacities and circumstances of citizens within a given society.

This moral grounding of political stability, in turn, makes possible what Rawls describes as “the deepest and most reasonable basis of social unity available to us in a modern democracy” (1996, 391). Justice must therefore arise *from within* the citizenry – it cannot be imposed externally. This acknowledges the constraints of political reality, since justice-based stability is neither a flawless, static equilibrium nor a perfect balance of power but a dynamic process that remains open to negotiation and revision.

A critic might now ask: does favouring the least advantaged not presuppose a fixed moral truth? To address this concern, we must consider the epistemic dimension of justice's intrinsic value, that is, how we come to understand what makes justice valuable in the first place. This brings us to the Kantian idea Rawls draws on, namely that persons are ends in themselves and are therefore free and morally equal (1996, 280). Since justice concerns *relations* between persons (*ibid.*, 278–281), rather than their individual capacities, it requires mutual recognition and respect for individuals as persons, rather than for what they have or can contribute. Justice, in this view, entails unconditional moral regard, consistent with Rawls's use of the original position and the veil of ignorance as the basis for defining fairness.

From this perspective, the least advantaged are no more or less valuable than the more advantaged. Given the universality of moral equality, each person is to regard others as equally entitled to fair treatment and to the distribution of

(dis)advantages. There is, therefore, no moral justification for privileging individuals based on merit, status, or inherent qualities. The difference principle follows directly from the intrinsic value of persons – what Kant refers to as *Würde*, or human dignity (Korsgaard 1986) – and, by extension, from the principle of equal moral worth.

In summary, realists are correct in observing that Rawls places justice prior to non-moral political concepts such as legitimacy (Williams 2005; Sleat 2014). However, they err in suggesting that Rawlsian justice is wholly pre-political. While justice does serve as a moral foundation, the substantive content of what counts as just is developed within the political process itself. We now turn to a more detailed examination of how that process unfolds.

2.4 POLITICS

2.4.1 Politics as a fair system of cooperation

For Rawls, politics serves an instrumental function, since its aim is to organise systems of cooperation that allow people with different conceptions of the good life to reach an agreement all can accept. Importantly for the discussions of realism and agonism that follow, Rawls treats politics as instrumental only in its structure – it is a means to justice rather than an end in itself, but it is not instrumentalist in its substantive justification (see 2.2.1).

Politics, for Rawls, begins with the view that society is a “fair system of cooperation” (1996, 15–22). However, fairness is not intrinsic to society; it must be actively established. And this is precisely the role of politics, namely to create a fair system in which conflicting views of the good can coexist. This commitment to fairness is based on the idea that all persons have equal moral standing, as discussed in section 2.5.1, and therefore deserve equal rights and opportunities. The two main principles of fairness – the equal basic liberties principle and the difference principle – provide a stable basis for political cooperation. In this sense, fairness is not a utopian ideal but something that can be built into political institutions and is necessary for a just society to function.

Thus, what is constructed is the institutional arrangement and the content of the political conception – not the order of values itself, which Rawls’s political constructivism only represents (1996, 95). The purpose of politics, then, is to construct the fairness of the procedures that govern social cooperation. Several assumptions underpin Rawls’s conception of politics.

First, politics is understood as a “system of cooperation, over time, from one generation to the next” (1996, 14). This transgenerational cooperation is central to Rawls’s theory, setting it apart from alternative perspectives that conceive of politics in synchronous terms, such as coordinative (communitarian), adversarial (agonistic), legitimacy-seeking (moderate realist), or power-centred (radical realist). Politics, in Rawls’s view, must create and maintain the conditions necessary for cooperation, which presupposes that political actors regard cooperation as the primary mode of engagement. However, Rawls’s theory, alongside this assumption, is idealised and does not reflect real-world circumstances, where individuals often differ not only in their conceptions of the moral good but also in their views on how collective life should be organised.

Second, Rawls holds that society is inevitably deeply diverse and therefore, in its pre-political state, disordered. Politics is thus tasked with achieving a “well-ordered” society (1996), one that is both stable and just. This diversity arises from our “different conceptions of the world” (*ibid.*, 57). Rawls argues persuasively that expecting everyone to adhere to a single, comprehensive philosophical, religious, or ethical worldview is unrealistic:

“Religious and philosophical doctrines express views of the world and of our life with one another, severally and collectively, as a whole. Our individual and associative points of view, intellectual affinities, and affective attachments, are too diverse, especially in a free society, to enable those doctrines to serve as the basis of lasting and reasoned political agreement” (1996, 57).

Rawls sees this foundational diversity as a permanent feature of free societies, and argues that political institutions must be designed to accommodate it (1996, xviii). He maintains that unity among worldviews is neither feasible nor desirable, as the justification for comprehensive doctrines is only convincing to those who already subscribe to them (*ibid.*, xxi). Consequently, external critiques of such doctrines are viewed as ineffective, making fundamental value disagreement unavoidable. Seeking consensus at this level would therefore not only be futile but also compromise freedom of thought, one of the central values of political liberalism.

Third, Rawls contends that only reasonable pluralism can be accommodated within a constitutional democratic regime. Unreasonable doctrines that do not conform to the framework of justice as fairness must be rejected, as they undermine the possibility of stability “for the right reasons” (*ibid.*, xix, lvii, 391). Reasonableness involves the capacity to distinguish between personal and political domains – between one’s deeply held commitments, closely tied to identity and well-being, and the principles required for cooperation in a

pluralistic society. In the personal sphere, self-respect requires that one's commitments be acknowledged and respected by others. Accordingly, the distinction between the private and the political must itself be part of a shared political understanding.

For Rawls, reasonableness is thus based on a willingness to consider perspectives other than one's own. The political realm is therefore characterised by voluntary participation and the absence of coercion – people take part in political life not because they have no other choice, but because they share a commitment to cooperation. Rawls writes:

“The first basic aspect of the reasonable, then, is the willingness to propose fair terms of cooperation and to abide by them provided others do. The second basic aspect, as I review now, is the willingness to recognize the burdens of judgment and to accept their consequences for the use of public reason in directing the legitimate exercise of political power in a constitutional regime” (1996, 54).

In this context, active willingness to participate in cooperative efforts is essential for effective political functioning. Passive cooperation – such as non-resistance – is insufficient, and cooperation under coercion is even more inadequate. Moreover, the idea of the *burdens of judgement* entails that participation requires a responsibility to critically assess not only our own claims but also those of others, taking into account the broader context. In other words, the political sphere must comprise responsible judges, since establishing common rules depends on all affected parties articulating their views on fair terms, engaging in self-reflection, and evaluating the truth and relevance of their claims (Rawls 1996, 55). The burdens of judgement lead us to recognise that our own reasonable comprehensive doctrines cannot be regarded as a priori superior to others, since no epistemic criterion exists by which competing reasonable doctrines can be conclusively adjudicated. As such, there is no compelling external reason for individuals to adopt different doctrines, only those that arise from within. Political cooperation is therefore grounded in the obligation to justify one's own reasons and the commitment to interpret and understand the reasons of others.

Fourth, the success of the political process is measured by whether the people involved can reach an agreement on shared political principles. Politics therefore has a specific goal and can, in principle, come to an end, since once that goal has been reached through fair and legitimate procedures, the issue no longer needs to be debated. In a diverse society, people do not need to agree on their moral beliefs; they merely need to agree on basic political principles concerning fair procedures (1996, 58). Importantly, collective life must be cultivated from within society rather than imposed externally. While this supports the view that

comprehensive philosophical (including moral and religious) doctrines belong to the private sphere, thereby suggesting a strict divide between personal and political domains, Rawls resists drawing overly rigid boundaries between them. Instead, he acknowledges their fluidity: “comprehensive doctrines of most people are not fully comprehensive, and this allows scope for the development of an independent allegiance to the political conception that helps to bring about a consensus” (ibid., 168). He continues: “Gradually, as the success of political cooperation continues, citizens gain increasing trust and confidence in one another” (ibid.). Accordingly, the procedural fairness embedded in constitutional arrangements fosters a constructive interaction between the personal and political spheres.

Although consensus might seem to imply compromise, Rawls draws a clear distinction. Consensus is based on “a balance of reasons as seen within each citizen’s comprehensive doctrine” (1996, 169), whereas compromise suggests that individuals are forced to concede something (ibid.). Rawls does not demand that political actors abandon their foundational commitments to achieve stability by conceding ground to others. He rejects a view of politics as a bargaining game. In his view, compromise concerns the distribution of power, while consensus is about agreeing on the rules that govern how power is exercised and constrained (ibid., 171).

What, then, allows fundamentally different worldviews to reach such consensus? Rawls suggests that each comprehensive doctrine “is related to the political conception in a different way” (1996, 171). That is, each accepts the political conception for its own reasons, based on its own internal logic. While this statement is too broad to explain how consensus could emerge across all doctrines, Rawls does not attempt to trace this path in each case, recognising that doing so would be an immense task. What matters is that all reasonable doctrines must recognise a distinction – however subtle – between the public and private domains.

Fifth, and building on the above, Rawls’s project requires a separation between political philosophy and foundational philosophical questions, and by extension, between political and private spheres. Rawls writes: “To succeed in finding such a consensus, political philosophy must be, so far as possible, suitably independent of other parts of philosophy, especially from philosophy’s long-standing problems and controversies” (1996, 171–172). This does not mean that political philosophy must ignore or reject substantive commitments that fall outside consensus. It need not be “sceptical of religious, philosophical, and moral truth, or indifferent to their values” (ibid., 172). Rawls warns that such scepticism or indifference “would put political philosophy in opposition to numerous

comprehensive doctrines, and thus defeat from the outset its aim of achieving an overlapping consensus” (ibid., 150). In short, dismissing foundational doctrines would undermine the very possibility of consensus.

Here, Rawls seems to caution against throwing out the baby with the bathwater. He assumes that most doctrines share enough common ground to support a political conception of justice. He writes: “since no political agreement on those disputed questions can reasonably be expected, we turn instead to the fundamental ideas we seem to share through the public political culture. From these ideas, we try to work out a political conception of justice congruent with our considered convictions on due reflection” (1996, 149–150). He continues: “We appeal ... to a political conception of justice to distinguish between those questions that can be reasonably removed from the political agenda and those that cannot” (ibid., 150), further insisting on the need to respect citizens’ comprehensive views: “To apply the principles of toleration to philosophy itself is to leave to citizens themselves the questions of religion, philosophy, and morals in accordance with views they freely affirm” (ibid., 154). Reasonable and tolerant doctrines, Rawls argues, will set aside their perceived truths when they realise others cannot accept them, instead prioritising widely shared principles of justice. This toleration does not seek to regulate all aspects of a doctrine; for instance, it does not determine the legitimacy of specific religious beliefs (ibid., 13).

To summarise the fifth feature: Rawls’s distinction between the political and the non-political is not absolute. Yet even in this moderate form, there must be a principled rationale for maintaining such separation, and this rationale is grounded in toleration towards deeply held beliefs.

This brings us to the sixth and final feature of Rawls’s conception of politics. His proposal to separate political from other domains implies that consensus requires respect. Crucially, this is not respect for individuals’ beliefs as such, but for the persons who hold them. This distinction is important, as it enables one to reject a commitment without thereby rejecting the person.

Rawlsian consensus is built precisely on this kind of respect for persons. This follows from the idea of equal moral worth and helps explain why individuals would endorse principles of justice based on reciprocity (M. C. Nussbaum 2011, 17 fn. 32). Within this framework, others are not treated as *a priori* adversaries to justice. Their beliefs and commitments are understood as revisable, while the worth of persons is taken to be prior to the judgements they make. Individuals are therefore prevented from claiming moral superiority over one another.

Rawlsian respect is not rooted in altruism but is instead entirely non-hierarchical. Respecting another person's commitments means recognising what those commitments mean to that person. This respect is derivative rather than direct, since what is respected is not the content of a person's conception of the good, but their capacity to form and pursue one. This is why it does not matter here whether the truth of a commitment can be shared by others. Respect therefore does not require agreement or endorsement. Nor does it extend to beliefs that violate foundational ethical principles, which rest on the idea that persons should be treated as ends in themselves and that their beliefs are integral to their identity.

Altruism typically involves acting for the general good, sometimes at one's own expense. Rawlsian reciprocity, by contrast, rests on mutual and equal recognition. He writes: "reciprocity lies between the idea of impartiality, which is altruistic (being moved by the general good), and the idea of mutual advantage understood as everyone's being advantaged with respect to each person's present or expected future situation as things are" (1996, 16–17). Reciprocity inherits its sense of fairness from impartiality, without requiring people to make sacrifices, and it inherits its focus on mutual benefit from the idea of mutual advantage. However, it measures benefit against a standard of equality rather than desert, which means that there is no moral reason to give some people more because they have greater merit, higher status, or have contributed more, and reciprocity must not be read as proportioning benefits to what each has contributed.

Consider, for example, a tax system in which wealthier citizens contribute proportionally a larger share of their income. The arrangement is reciprocal not because the benefits each receives are proportionate to their contributions – they are not – but because its terms are ones that each citizen could reasonably accept as free and equal. If the greater tax burden were justified by the indirect benefits it secures for those who bear it – e.g., a more stable society or a healthier and better-educated citizenry – reciprocity would collapse into the very notion of (utilitarian) mutual advantage from which Rawls seeks to distinguish it. As a justificatory principle, reciprocity makes a more demanding claim, namely that the greater advantages of the better-off are permissible only insofar as they improve the position of the least advantaged. This follows from the principles that parties would choose in the original position, behind the veil of ignorance, and ultimately rests on Rawls's conception of citizens as possessing the two moral powers and, accordingly, as moral agents (see 2.5.1).

Rawls defines reciprocity as a principle according to which "all who are engaged in cooperation and who do their part as the rules and procedures require, are to benefit in an appropriate way, as assessed by a suitable benchmark of

comparison” (1996, 16). Reciprocity is the appropriate standard for justice because it rests on the equal worth of all citizens, and in this way it creates a moral bond between citizens and the laws they follow. Citizens obey the law not simply because doing so serves their own interests but because they see its rules as reasonable once they take both their own interests and those of others into account. This is not a technical calculation, but a moral judgement grounded in the idea that persons, as moral agents, are capable of ethical reflection and deserve equal consideration.

Ultimately, Rawls’s conception of politics rests on a mode of reasoning that seeks to avoid impasses rather than establish the supremacy of any particular doctrine. For Rawls, the diversity of comprehensive doctrines is not a defect but the natural outcome of free institutions in a liberal society: “[doctrines] are not simply the upshot of self- and class interests, or of people’s understandable tendency to view the political world from a limited standpoint, but the work of free practical reason within the framework of free institutions” (1996, 36). Accordingly, reasonable pluralism does not seek to determine the truth or falsity of comprehensive moral doctrines by asking, ‘What is justice?’ Instead, it asks, ‘What is just?’ and ‘What political arrangements can free and equal citizens reasonably endorse as just?’ These questions presuppose a conception of justice to be constructed through political agreement rather than a moral truth to be discovered. What Rawls leaves unaddressed is not what justice is or what is just, but how political actors come to see what matters and how its meaning is formed in the first place. It is this question that chapter 6 takes up in developing the ethos of creative valuation.

2.4.2 Liberal legitimacy

The discussion of politics established that reasonable persons, reasoning conscientiously, nonetheless diverge, and that no doctrine’s account of moral truth is publicly available to arbitrate between them. This invites a question for legitimacy: if no doctrine owns the truth about justice, on what basis may anyone be coerced in its name? For Rawls, legitimacy is essential because it ensures that political power is exercised in a way that is *morally justifiable* to all citizens as free and equal individuals. It is not merely about legal authority or procedural compliance, but about offering *sincere* reasons that others can reasonably accept, rooted in the moral duty of civility. For Rawls, legitimacy sustains a just political order by treating political power as a collective expression of the citizenry, not as a tool of domination: “while political power is always coercive power, in a constitutional regime it is the power of the public, that is, the power of free and equal citizens as a collective body” (1996, 68). Thus, rather than attempting to

abolish the coercive character of political power, Rawls seeks to reconstitute it as an expression of the collective will of free and equal citizens.

In the broadest sense, legitimacy concerns the relationship between those who wield power (the state) and the citizens. According to liberal theories, the coercive actions of the state are considered legitimate only insofar as they are voluntarily accepted by those subject to such coercion. This acceptance obliges the state to offer justifications for its actions. Political authority, therefore, fundamentally depends on the consent of the governed. Such consent places limits on state power, which, if left unconstrained, could otherwise tend towards absolutism. In essence, consent underpins legitimate political authority, which is granted when the state sufficiently justifies its use of coercive power. To emphasise, Rawls's criterion is not consent, whether actual or hypothetical, but reasonable acceptability, that is, what citizens "may reasonably be expected to endorse" (1996, 137), so that legitimacy turns on the character of the reasons offered, not on whether assent has in fact been given.

In Rawls's framework, the link between consent and justification is captured in the criterion of reciprocity within public reason, itself grounded in the idea of sincerity:

"Our exercise of political power is proper only when we sincerely believe that the reasons we offer for our political action may reasonably be accepted by other citizens as a justification for those actions" (1996, xlvii).

The emphasis on sincerity suggests that legitimacy is not simply a transaction in which each side does just enough to reach an agreement. Instead, it involves an ongoing dialogue, where those in positions of power genuinely take citizens' reasonable interests into account when explaining and justifying their decisions. These explanations should not be manipulative, but should show respect for citizens' ability to make moral and informed judgements. Quong refers to this kind of respect as "civic friendship" (2011, 265), which recognises others as rational agents capable of moral reasoning. As he explains:

"If we acted insincerely towards other citizens, if we offered arguments we believed to be invalid, or which we believed others had no good reason to accept, we would fail to respect their status as citizens who can understand and respond to moral reasons, and are owed justifications for the rules that regulate social cooperation" (2011, 266).

Although Rawls does not explicitly refer to it, sincerity also implies epistemic humility – namely, a willingness to revise one's own commitments if they are shown, through political dialogue, to be unreasonable.

Although the demand for sincerity may at first seem idealistic, it becomes easier to understand once we see what Rawls wants to stress in his account of legitimacy, namely that politics is not simply a matter of power and coercion. Instead, he emphasises the ethical side of political reasoning and distinguishes it from the strategic calculations that often shape political practice. For Rawls, people have a moral nature, and this moral dimension remains important in political life. He writes:

“We must start with the assumption that a reasonably just political society is possible, and for it to be possible, human beings must have a moral nature, not of course a perfect such nature, yet one that can understand, act on, and be sufficiently moved by a reasonable political conception of right and justice to support a society guided by its ideals and principles” (1996, lxii).

Politics, in this view, is not the central purpose of life but a means of creating a just and morally acceptable framework for collective living. Political power itself is not central; rather, it is the *powers of the citizen* (1996, 47–88) – their moral and epistemic capacities, as well as their capacity for mutual respect and reciprocity – that lie at the heart of political life. The point of politics is not to legitimise power struggles, but to respect individuals’ freedom to pursue their own lives within a system of moral equality. Rawls articulates this vividly:

“If a reasonably just society that subordinates power to its aims is not possible and people are largely amoral, if not incurably cynical and self-centered, one might ask with Kant whether it is worthwhile for human beings to live on the earth?” (1996, lxii).

Accordingly, the legitimacy of political authority hinges on public reason (1996, 212–254). For political power to be just, it must be exercised in accordance with principles that all citizens can reasonably accept – principles expressed through reasons that are publicly accessible, regardless of citizens’ own comprehensive doctrines. The question of legitimacy is framed as follows: “In light of which principles and ideals must we exercise [coercive political] power for our actions to be justifiable to others as free and equal?” (1996, 217). His answer is:

“Our exercise of political power is proper and hence justifiable only when it is exercised in accordance with a constitution the essentials of which all citizens may reasonably be expected to endorse in the light of principles and ideals acceptable to them as reasonable and rational” (1996, 217).

Rawls does not conceive of political power and citizens as separate or opposing forces. Instead, he describes political power as “the power of the public, that is, of

free and equal citizens as a collective body” (1996, 68). The relationship between state and citizen is not antagonistic, as between dominator and dominated; political power is not inherently an instrument of domination. Rather, it is a legitimate form of authority that is shared and exercised collectively. The need for legitimacy is therefore not merely formal or procedural (for example, to maintain institutional balance), but also moral, based on reciprocity and on the recognition that those who exercise power are part of the same political community as those they govern, even if they have roles that carry particular responsibilities. As Rawls writes:

“And since the exercise of political power itself must be legitimate, the ideal of citizenship imposes a moral, not a legal, duty – the duty of civility – to be able to explain to one another, on those fundamental questions, how the principles and policies they advocate and vote for can be supported by the political values of public reason” (ibid., 217).

The *moral duty of civility* reflects the principle that all citizens, in their political roles, must respect others’ reasonable doctrines and support only those policies that they believe all reasonable people could endorse. This duty implies an epistemic constraint whereby political justifications cannot appeal to the moral truths of any particular doctrine, however benevolent those truths may seem, given that such truths give rise to deep disagreement because no doctrine is in a position to claim them as its own.

Because moral truth in the political domain is neither absent nor simply plural, it can be viewed as *unowned*, in the sense that no single doctrine can claim it. Rawls does not deny that a particular doctrine may be true. Rather, he argues that no one can publicly establish *which* doctrine is true. Citizens may recognise one another’s doctrines as reasonable without accepting them all as true, and there is no shared public standard for deciding between competing claims to truth (ibid., 128). If one doctrine could establish itself as the owner of moral truth, it would also gain a claim to rule, but if that possibility is removed, political authority rests instead on the equal standing of citizens. Thus, political power belongs to the collective body not because it possesses the truth, but because no one can claim public authority over it.

The authority of the collective body is therefore procedural rather than epistemic, and its legitimacy depends on the fairness of the terms under which power is exercised, not on the truth of the moral convictions underlying political decisions. A majority is not right simply because it is a majority, and this is why the duty of civility binds citizens even when they are convinced that their own moral beliefs are true.

The unownedness of moral truth is therefore not a problem that the collective body must overcome but the condition that makes collective political authority possible in the first place. If one of the reasonable doctrines within the overlapping consensus is in fact true, then the political conception it supports is also correct, and all reasonable doctrines endorse the right conception of political justice, even if they reach it for different reasons (*ibid.*, 128). Politically speaking, if any one citizen is right, then all citizens are right.

The idea that civility is framed as a moral rather than a legal duty highlights two further points. First, political decisions must be grounded in moral considerations. Second, while legal duties are externally imposed, moral duties arise from respect for citizens' moral and epistemic capacities, including their ability to distinguish between the civic and the personal. As Rawls observes, this duty entails being "guided by the criterion of reciprocity whenever constitutional essentials and matters of basic justice are at stake" (1996, lv). As discussed earlier, reciprocity is rooted in respect for individuals as intrinsically valuable persons. The duty of civility is therefore a moral obligation because it flows from the recognition of human dignity – a principle central to the Kantian tradition on which Rawls builds. Here, justice is not simply about balancing interests, but about protecting the dignity and moral worth of individuals. Rawls underscores this point with this statement: "If justice perishes, then it is no longer worthwhile for men to live upon the earth" (*ibid.*, lxii fn. 40).

In summary, Rawlsian legitimacy is not a competitive struggle in which justification serves merely to secure power, however well-intentioned. It consists in exercising political power on terms that all citizens may reasonably be expected to endorse, and these terms are grounded in mutual moral respect and the recognition of citizens as free and equal. These terms are necessary because no one can claim to own or have exclusive authority over moral truth, and thus, no doctrine has a legitimate claim to political power. However, legitimacy remains distinct from justice, since legitimacy concerns the manner in which power is exercised, whereas justice is about the fairness of the institutions and laws that result from that exercise of power, and thus it is possible for a political order to be legitimate without being fully just. And although it rests on a moral conception of politics, it does not collapse into moralism, as it seeks no consensus on comprehensive moral principles – these belong to citizens' own (private) doctrines – but establishes only the conditions under which citizens who hold different ones may share a common political framework.

2.5 POLITICAL ACTORS

Envisaging a system of principles for a just society presupposes a particular conception of the political actor, and depends on how we understand individuals both as autonomous agents and as participants in the political sphere. What, then, are the qualities that define Rawlsian political actors – those necessary for establishing a just political order through a fair deliberative process?

Arguably, the most salient qualities of such actors concern the way in which they make decisions. According to Rawls, political actors are expected to make decisions not solely on the basis of pure rationality, but within the framework of public reason. He draws a clear distinction between the rational and the reasonable. Rationality refers to reasoning grounded in factual knowledge, which “can be modelled by mathematical assumptions” and is therefore predictable (Rawls 2003, 81). Reasonableness, by contrast, is “a moral idea involving moral sensibility” (ibid., 7), whereby individuals acknowledge moral constraints of reciprocity by recognising others as moral equals. This recognition forms the foundation for “fair terms of social cooperation” (Rawls 1996, xliv), wherein individuals agree to abide by shared principles “even at the cost of their own interests in particular situations, provided that others also accept those terms” (ibid.).

Let us now turn to what this *moral sensibility* of reasonableness entails.

2.5.1 Moral powers

For Rawls, being a reasonable agent involves possessing two moral powers (1985, 233–234; 1996, 103–104), which together constitute the “basis of moral motivation in the person” (1996, 81) and provide individuals with reasons to seek political consensus.

The first of these is the *capacity for a conception of the good*, defined as “the capacity to form, to revise, and rationally to pursue a conception of one’s rational advantage, or good”, where ‘good’ refers to that which is “valuable in human life” (1985, 233). This includes both intrinsically valuable aspirations and the social and interpersonal attachments that contribute to a sense of human flourishing.

This capacity can be understood as having two dimensions. First, it involves valuing something for its own sake, rather than merely as a means to an end. Second, intrinsic values often have social origins, as their significance is shaped through relationships with others. Accordingly, the capacity to conceive of the moral good requires recognition that a genuinely ‘human’ life cannot be lived in isolation, since the notion of the good is inherently relational in its development. While the good possesses intrinsic worth, its existence is inextricably linked to its

relational nature, as the concept of a *human person* implies a social entity (Baker 2015).

Consequently, rational individuals pursue their conception of the good through a process of implicit or explicit negotiation with others – that is, by seeking to become “fully cooperating members of society” (1996, 81). This cooperative process provides the rationale for seeking consensus, as the very meaning of the good is embedded in social interaction. Even if a person’s doctrine is not intended to persuade others, it must nevertheless be articulated in order to remain intelligible and relevant. In this sense, the human being becomes a political actor by virtue of their commitment to a doctrine, which in turn grounds their capacity for a conception of the good. Each citizen adheres to a comprehensive doctrine that includes its own “theory of truth [and] status of values” (1996, 150). Without such a doctrine, an individual would lack the moral resources required for meaningful political participation. Crucially, this doctrine must rest upon a moral foundation that affirms the moral equality of persons. Both the belief in moral superiority – which will inevitably shape political life if left unchallenged – and the complete absence of morality would indicate the lack of a genuine conception of the good.

The second moral power is the *capacity for a sense of justice*, defined as “the capacity to understand, to apply, and to act from the public conception of justice which characterises the fair terms of social cooperation” (1985, 233). At first glance, this might appear to be an innate human quality; however, Rawls does not treat it as such. Rather, he regards it as a capacity developed within the political process, particularly through the procedural structures of a well-ordered society – most notably the principles of symmetry (equality) and the veil of ignorance, which ensure impartiality. Citizens cultivate this capacity through participation in public political reasoning, guided by a shared and publicly recognised conception of justice. This process also fosters reflexivity, encouraging individuals to consider how mutually accepted principles shape both their self-understanding and their motivation to act justly. A Rawlsian well-ordered society can therefore maintain long-term stability if its citizens treat the sense of justice as something that requires continuous attention and development. As Rawls notes: “those who grow up under just basic institutions acquire a sense of justice and a reasoned allegiance to those institutions sufficient to render them stable” (1996, 142).

2.5.2 Epistemic powers

Additionally, Rawls identifies two interconnected features of political agents that contribute to the epistemics of consensus-seeking in a pluralistic society. The first

is the intellectual capacity for judgement, thought, and inference; the second is the recognition of the *burdens of judgement*, which “limit what can be justified to others and affirm only reasonable comprehensive doctrines” (1996, 81). Within this framework, individuals are expected to possess the cognitive abilities necessary for meaningful engagement with others – particularly through critical thinking and reflexive participation in public deliberation. This involves openness to the plurality of political truths and a willingness to collaboratively construct shared understandings of justice.

Crucially, this includes the capacity to reason about the nature of judgement itself – specifically, to discern what kinds of claims and reasoning are appropriate within public political discourse. The primary concern, then, is not the ability to persuade others, but the capacity to distinguish between claims that are publicly justifiable and those that reflect private convictions. As previously discussed, the basis for such reasoning lies in mutual respect and tolerance for other reasonable doctrines. However, this respect – grounded in the equal inherent moral worth of all individuals – is not unconditional; it requires a principled evaluation of which doctrines are reasonable and which are not. On this basis, Rawls excludes proponents of unreasonable doctrines (1996, 39) – those that reject the moral equality of persons – because such doctrines fail to meet the standards of respect and reciprocity required for political justification.

For Rawls, the epistemic dimension of the burdens of judgement (1996, 54–58) – that is, understanding the sources of disagreement – is a rational endeavour rooted in the recognition that reasonable pluralism arises not from irrationality, but from the complexity of human reasoning. Unreasonable views result from a failure to engage with this complexity, reflecting a lack of epistemic commitment and producing flawed reasoning marked by “prejudice and bias, self- and group interest, blindness and wilfulness” (ibid., 56). Rawls identifies several sources of reasonable disagreement (ibid., 56–57), including conflicting empirical evidence, divergent value priorities, conceptual vagueness, the diversity of personal experience, the complexity of moral ends, and practical limitations in balancing competing values.

These factors all pertain to epistemic rationality, demonstrating how abstract principles interact with real-world conditions. This perspective does not deny the existence of poor epistemic practices in political life – that is, disagreements that stem from a refusal or inability to engage seriously with the views of others. Rather, it suggests that unreasonable judgement emerges from a lack of epistemic responsibility, which is essential for understanding the empirical and normative complexity underlying pluralism.

However, Rawls does not fully address how to respond to the presence of *unreasonable doctrines*, even though pluralism inevitably includes not only reasonable but at times also hostile or exclusionary systems of belief. If respect is granted only to reasonable citizens, does this not suggest that it is not the intrinsic moral worth of persons that underlies respect but the nature of their commitments? Rawls does not offer concrete examples of unreasonable doctrines, but one might speculate that this omission stems from the assumption that overtly harmful ideologies (e.g., religious extremism) do not qualify as comprehensive doctrines in the Rawlsian sense. That is, they lack the depth and coherence required to function as holistic moral frameworks, and are therefore considered ideologically ‘thin’. On this interpretation, Rawls’s commitment to the intrinsic moral worth of persons, articulated through reasonable comprehensive doctrines, remains intact. While such unreasonable outliers may be rare, the use of coercive measures against them may be justified on practical grounds, particularly to prevent significant harm.

2.5.3 Emotional powers, or the lack thereof

With the moral and epistemic powers of Rawlsian political actors outlined, one might question whether such rational, duty-based capacities are sufficient to secure consensus. Rawls appears to believe they are, as he does not explicitly address the role of emotive or affective capacities in fostering respect and reciprocity. He refers only in passing to ‘sentiments’, ‘attachments’, ‘devotions’, and ‘affections’ that characterise our interpersonal and collective lives, which we subsequently come to value as part of the good. For Rawls, then, emotive elements are integral to human relations and should be acknowledged by rational agents, but emotions are not required to play a role in consensus-building. In other words, while we should recognise others’ emotive attachments, Rawls maintains that justification of judgements must rely on reason alone.

Some critics, such as Sibyl Schwarzenbach, argue that Rawls diminishes the conception of the political person by abstracting them too far from the concrete and particular dilemmas of ethical and political decision-making. On this view, the capacity for a sense of justice is developed not solely through abstract principles but also through engagement with real-world moral complexities and social contexts. Schwarzenbach proposes a third moral power: an “abstract capacity for *philia* (friendship)”, which enables individuals “to note and interpret the particular needs of fellow citizens (conceived as moral equals) and to respond appropriately with good will and practical care in the concrete” (Schwarzenbach 2009, 116).

Rawls's shortcoming, then, according to Schwarzenbach, lies in his portrayal of political actors as (almost entirely) rational beings. In reality, tolerance and respect often rely on the ability to *empathise* with others and to understand their perspectives. This suggests that fairness – understood as the equal exchange of reasoned approvals – may be insufficient to establish mutually acceptable terms. A purely rational model of consensus-seeking neglects the essential role of emotions in social relations, thereby reducing political agency to a merely functional role that individuals are expected to perform.

From Rawls's perspective, this exclusion of emotion seems warranted, since achieving reflective equilibrium requires a clear distinction between personhood and the political role. However, critics may argue that this approach fails to account for the lived experiences of others, potentially leading to impasses in efforts towards social progress. Consequently, some theorists – particularly feminist scholars (e.g., Abbey 2015) – argue that Rawlsian contractualism tends to reinforce the status quo. This critique holds that rational consensus among differing doctrines cannot extend beyond what is already shared and familiar. By contrast, genuine progress requires not only challenging one's own doctrine but also interrogating dominant and entrenched truths. Without such critical reflection, the emergence of new ideas becomes unlikely, if not impossible. On this view, progress depends on the capacity to introduce novel (and often external) perspectives. Acknowledging the emotive dimensions of human interaction can therefore expand the boundaries of consideration, thereby enabling transformative change.

One might reasonably argue that it is precisely the abstraction and role-taking central to Rawlsian theory that make it internally coherent. The original position, the veil of ignorance, and the principle of reciprocity are all impersonal mechanisms foundational to Rawls's theory of justice. Within this framework, the function of respect and tolerance is to bracket personal doctrines, based on the assumption that emotions could blur the line between the personal and the political.

In Rawlsian liberalism, political actors' personal identities are to be separated from the political sphere. They are expected to conceive of themselves as public and impartial agents, tasked with resolving political questions according to an abstract and shared set of rules. The cost of this, however, may be a form of ethical detachment, where individuals feel less accountable for interpersonal responsibilities. For example, it can lead to negligence when someone in an official role puts following the rules ahead of a person's real needs, and in a similar way it can lead to politico-moral corruption, such as partisan voting,

when personal convictions clash with party directives and the resulting actions serve the party instead of the public good.

Moreover, a critic might observe that excluding emotions from political reasoning also excludes the motivations essential for seeking mutual agreement. Rawls, by contrast, identifies just and reasonable stability as the central purpose of politics. Yet in the context of social advancement, the drive to participate in politics cannot be motivated solely by justice. It also stems from a desire to contribute to positive change in a broader sense and to participate in the creation of something new. This sense of contributing to a greater collective purpose may be far more fulfilling – and politically motivating – than merely improving one’s own circumstances or even those of the least advantaged.

Schwarzenbach similarly contends that the responsibility of a political actor does not end with reaching agreement on how to coexist – that is, maintaining a well-ordered society. Rather, it entails an ethical conception of political agency committed to enhancing the norms, rules, and standards of communal life. She argues: “Not only the construction, but the maintenance, reproduction, and improvement of such institutions presuppose a fundamental responsibility on the part of each citizen to care for and (ethically) befriend fellow citizens” (2009, 118–119), further warning that

“If this third moral power [capacity for *philia*] is not explicitly recognised, encouraged, and publicly reinforced, moreover, then the social institutions grounded on it will likewise begin to lose their viability. Indeed, in the present atmosphere of individual license, hectic production, and ‘privatization,’ the important public institutions mentioned here are increasingly being called into question” (ibid., 119).

As we construct increasingly impersonal systems, norms, and institutions to organise our shared existence, we risk weakening our capacity to recognise fellow citizens as *human beings*. This may reduce our willingness to offer direct support, as we come to rely on systems to discharge our obligations. For instance, when drug abuse is viewed solely as a structural issue associated with one’s place in society, rather than something individuals can confront through autonomy and personal responsibility, we risk undermining agency and hindering the potential for self-help.

2.6 CRITIQUE OF RAWLS

2.6.1 (In)Feasibility

Critics frequently contend that Rawls's ideal theory of achieving stability through overlapping consensus is both unfeasible and insufficiently ambitious in promoting social change. The claim that ideal theories fail to provide adequate action-guidance is, of course, a familiar one (e.g., Farrelly 2007; Stemplowska 2008; Wiens 2015).

This section examines the kind of feasibility relevant to evaluating Rawls's ideal theory of political liberalism. At first sight, Rawls might appear to develop a theory intended to guide action under non-ideal conditions rather than merely describe an unattainable ideal. Chahboun (2019, 551), for example, argues that unlike some ideal theories, such as Cohen's (1989) luck egalitarianism, which is not intended to be fully realised, Rawls's justice-oriented political liberalism does aspire to implementation, given its emphasis on procedural and institutional dimensions⁷. Before assessing this claim, however, it is important to recall that Rawls's theory is not deontological in the sense often attributed to it. Although he formulates the principles of political justice in terms of rights (1996, 109 fn. 15), these rights are not derived from a prior conception of the good (see 2.2.2), nor do they impose categorical moral duties on citizens independently of the political process.

His consensus-seeking principles are accordingly grounded in contractarian reasoning (Southwood 2009; Cudd and Eftekhari 2025), which implies that agreement is to be enacted rather than merely stated. But is it appropriate to assess action-guidance solely in terms of practical applicability to non-ideal circumstances? Put differently, should Rawls's theory be read as a guide for pursuing justice within existing societies, or might it aim at a different kind of feasibility altogether?

Consider the feasibility of principles along two dimensions of action-guidance: usefulness and meaningfulness. *Usefulness* refers to the functional aspect of principles, their capacity to delineate the most efficient and practicable means of achieving a given end, which entails adaptability to shifting circumstances and a strong connection to empirical evidence, that is, to a reliable account of what is the case (Dowding 2020; Perez 2020, 2024). *Meaningfulness*, by contrast, concerns the justificatory dimension, namely the reasons principles give for why and how a particular action advances the goals they are meant to serve. In the context of Rawlsian justice, a principle may be considered *useful* if it facilitates overlapping

⁷ It should be noted, however, that Rawls himself regarded his theory as purely a hypothetical construct (see 2.2.2).

consensus with minimal normative loss (given that pluralist contexts necessarily entail some compromise), and *meaningful* if it explains why those actions contribute to justice.

In this way, action-guidance involves considering what the principle is intended to represent or direct in its application – is it primarily concerned with function or meaning? Rawls draws an analogy between the operation of the sense of justice and our intuitive use of language, comparing it to our ‘sense of grammaticalness’ in our native tongue ([1971] 1999, 41–42). As native speakers, we speak fluently without conscious awareness of grammatical rules, whereas in learning a foreign language, we must grasp the grammar alongside memorising vocabulary. In a similar vein, Rawls argues that our innate sense of justice operates according to principles akin to grammatical rules, enabling mutual intelligibility and facilitating agreement on the substantive content of justice.

Extending this linguistic analogy, one might say that the principles of ideal action-guidance delineate the style of ethical expression, much as, for example, distinctions between sarcasm, irony, and mockery, or between poetry and theatrical performance, delineate different styles of expression. Additionally, these principles provide a basis for explaining why a particular style is preferable, which, in the case of Rawls, is stability for the right reasons (Rawls 1996, 391).

Put differently, in seeking consensus on justice, ideal principles articulate the stylistic features of the ethical framework being adopted, while leaving moral agents, as political actors, to determine the specific substantive content of justice. In this respect, ideal principles are *useful* in guiding ethical reasoning by ensuring coherence and consistency in our judgements, and *meaningful* in offering interpretive frameworks for understanding justice – akin to helping us distinguish, for instance, between a cynical and a sarcastic remark, or a poem and a play.

By contrast, non-ideal principles operate by determining whether a substantive content of specific action is permissible in light of present circumstances. This is because our understanding of non-ideal contexts must be derived from empirical knowledge, which pertains directly to substantive content. As such, the normative principles of non-ideal theory tell us what we should do ethically in real-world situations, while Rawls’s ideal theory gives us a framework for thinking about ethics and comparing different ways of reasoning about what is right.

Thus, if we assess feasibility in terms of grounding action-guidance in real-world conditions (i.e., from a non-ideal perspective), we evaluate principles based on whether they provide sufficient empirical and substantive information to guide

action. Conversely, if feasibility is conceived in terms of framing the constraints that define the register or style through which we approach the political domain (i.e., from an ideal perspective), then we evaluate principles according to the degree to which they offer conceptual coherence. Rawls's justice-oriented political liberalism aligns with the latter approach, since it does not try to account fully for empirical complexity. Instead, it works within the normative framework of liberalism, and its ideal principles are meant to set out the structural requirements of the liberal register, not, for example, to guide a transition from theocracy to liberal democracy.

Accordingly, if we were to construct a theory to support such a transition – from, say, a theocratic to a liberal state – we would need to incorporate practical considerations relevant to that transformation. Rawls's theory, however, does not pursue this goal. It begins *ex nihilo*, assuming that individuals are not situated within an unjust political system but are instead in a pre-political condition – neither citizens of an existing state nor inheritors of prior political status. The theory is not concerned with designing an ideal conception of justice per se, but with identifying the conditions under which stability can be achieved amidst deep moral pluralism, enabling justice to emerge from within that same diversity and thereby securing stability for the right reasons.

Leader Maynard (2024, 6 fn. 14) reaches the same conclusion by his own route, arguing that Rawls's idealisation is primarily heuristic rather than telic. Three features of the theory support this, namely its emphasis on reflective equilibrium, its prioritisation of coherence over moral truth (Rawls 1974), and its use of the original position not to reach a logically necessary conclusion but to make a conjecture about which conception of justice the parties would choose ([1971] 1999, 30).

In sum, the claim that political liberalism is infeasible as a means of guiding responses to empirical contingencies is accurate, but reductive, since it holds Rawls to a purpose his theory never assumed. To use a further analogy, he does not prescribe the melody that makes up justice but supplies the elements from which one might be composed, such as scales, chords, intervals, rhythm, harmonic structures, and compositional forms. The composing is left to moral agents working within a shared ethical register.

2.6.2 Power and the status quo

Rawls has been criticised for offering no adequate account of how existing inequalities and forms of oppression are sustained, on the grounds that his political conception of justice fails to confront power as the force underpinning prevailing social structures and, in doing so, perpetuates the liberal capitalist

status quo (Geuss 2008; Young 1990). Critics writing from traditions of social critique argue that, without attending to power, it is impossible to address inequality effectively, since this requires exposing the mechanisms through which dominant groups legitimise institutions and practices that serve their own interests. This criticism finds some support in Rawls's writing, since he does not treat power as a concept in need of analysis (see also Geuss 2008, 90), and the term does not even appear in the index of *Political Liberalism* (1996). Let us now consider this in more detail.

The most sustained critique comes from feminist scholarship (for an overview, see Nussbaum 2002), which identifies both methodological and substantive shortcomings in Rawls's account, amounting to a failure to address historically entrenched and gendered asymmetries of power. Marion Smiley argues that Rawls's methodological reliance on abstraction and individuation sidelines the values central to the ethics of care and relational approaches, thereby inadvertently reinforcing patriarchal social structures (2004, 1601). Moreover, she contends that the original position encourages agents to reason in terms of interests, albeit intrinsically rather than instrumentally valued. At the substantive level, Smiley argues that the "veil of ignorance" is ill-suited to challenging patriarchal family relations, since it brackets agents' social identities and, in doing so, undermines their moral agency and personal integrity (*ibid.*). The broader question raised by this critique is whether the abstraction characteristic of ideal theory and contractarian reasoning commits Rawls to a conception of political justification grounded in the exchange of reasons between self-interested agents.

These objections share a premise about what power is and where it resides. The conception of power used in social critique is broadly Hobbesian, in that power is seen as present in all social relations and as a basic human drive to satisfy individual desires (Grcic 2007, 372). On this view, relations between people, whether direct or shaped by norms and institutions, are at bottom power relations, in which individuals form and satisfy their needs by exercising power over others. Politics accordingly becomes a matter of managing the balance of power through contractual arrangements in which power is exchanged for goods or benefits (*ibid.*, 375), power itself functions as a bargaining tool among rational agents pursuing their interests, and justice is understood as strict adherence to those contracts (*ibid.*, 378). However, Rawls's framework does not seek to regulate power; rather, it gives us reason to reject power as the foundation of politics altogether, and thus to resist framing politics in Hobbesian terms in the first place.

The critique that charges Rawls with abstraction interprets the original position more narrowly than the text may justify. Smiley herself acknowledges that it

invites empathy (2004, 1607), and while Rawls does not explicitly refer to empathy (see also 2.5.3), neither does he present his theory solely in terms of rational choice. He explicitly rejects the idea that political agreement is a purely self-interested exercise:

“Citizens are reasonable when, viewing one another as free and equal in a system of social cooperation over generations, they are prepared to offer one another fair terms of social cooperation (defined by principles and ideals) and they agree to act on those terms, even at the cost of their own interests in particular situations, provided that others also accept those terms” (1996, xliv).

For Rawls, political justice is primarily a moral conception, in which consensus-seeking is grounded in intrinsic moral values rather than rational calculation, and the pursuit of self-interest is constrained by a sense of justice comprising duties of reciprocity, respect for others as free and equal, and the duty of civility. Political agents, as moral beings, are not atomistic but interdependent. As Okin observes:

“Those in the original position cannot think from the position of nobody, as is suggested by those critics who then conclude that Rawls’s theory depends on a ‘disembodied’ concept of the self. They must, rather, think from the perspective of *everybody*, in the sense of each in turn.” (1989, 101, emphasis added)

A strong status quo bias would indeed arise if the original position were understood solely as a mechanism for promoting self-interest *ex nihilo*, with individuals maximising personal gain on the basis of their actual capacities and converting luck-based advantages into power. But as Okin suggests, the difference principle and its maximin assumption emphasise interconnectedness rather than detachment, and thereby open the possibility of reconfiguring social arrangements and challenging inequalities, including those based on gender, by requiring that agents sustain the ‘perspective of everybody’ throughout public justification. The distinction between consensus and compromise (see 2.4.1) matters here too, since compromise involves negotiation driven by interests, whereas Rawls advocates consensus, in which morality acts as an epistemic constraint that filters interests and, with them, political concepts.

To address the critique that charges Rawls with neglect of power, we can reply that power is not so much absent from Rawls’s theory as relocated within it. Where Hobbesian contracts are underpinned by power serving self-interested ends, Rawls dissolves its role by embedding it in the “formal constraints” of justice ([1971] 1999, 130; cf. Grcic 2007, 390), constraints designed to apply across generations and to all moral persons, enabling individuals to prioritise moral

values and serving as the ultimate source of justification. Power is thereby transferred to these constraints, such that no one can claim ownership of it; to the extent that it is used to justify moral claims, Rawlsian power is *unowned* because, as discussed in 2.4.2, moral truth itself is unowned. The political person, not individual interest or relative advantage, lies at the heart of politics: “in a democracy, political power, which is always coercive power, is the power of the public, that is, of free and equal citizens as a collective body” (1996, 216). Politics, so conceived, is not the management of struggles for power but the empowering of individuals to make meaningful life choices within a framework securing moral equality. Rawls states the rejection of power-centred politics most directly in the closing paragraph of his introduction:

“The wars of this century with their extreme violence and increasing destructiveness, culminating in the manic evil of the Holocaust, raise in an acute way the question whether political relations must be governed by power and coercion alone. If a reasonably just society that subordinates power to its aims is not possible and people are largely amoral, if not incurably cynical and self-centered, one might ask with Kant whether it is worthwhile for human beings to live on the earth?” (1996, lxii)

Formal constraints, on this view, exist to articulate a shared system of publicly accessible principles, not to legitimise or regulate struggles for dominance. By refusing to ground politics in power, Rawls avoids the tendency of power-based relations towards rivalry and competition, which in extreme cases produce the “manic evil” he invokes (1996, lxii).

The disagreement, then, is not whether Rawls noticed power, but whether politics ought to be theorised in its terms at all. On the methodological charge, the critique seems weak, as Rawls’s framework deliberately tries to move beyond the status quo, both methodologically, by abstracting from existing social structures through moral constraints, and substantively, by setting aside the social positions and natural endowments that give some people an advantage over others, such as race, gender, or inherited talent, because these are arbitrary from a moral point of view.

Critics advocating radical social change, by contrast, make a case for bringing constructed identities to the forefront. But this comes at a cost, because although it can expose entrenched power dynamics, it would require openly relying on moral claims that are context-specific and subjectively grounded. For instance, using gender identity as a basis for reform through, e.g., quotas, may itself prove exclusionary, marginalising those who conform to neither binary nor non-binary categories, or who reject the category of gender altogether. Rawls’s original

position presupposes no identity-based assumptions at all; its critique of the status quo stems precisely from the deliberate exclusion of all endowments, natural, cultural, or self-assumed, allowing a shared conception of justice to emerge through the interaction of morally equal and free agents.

One version of the critique of Rawls's relationship with power nevertheless survives, namely Geuss's objection that Rawls does not simply neglect the workings of power, but that abstracting from power is itself an ideological move, presenting a historically contingent liberal settlement as the product of impartial reason (2008, 88–90). In other words, while for Rawls the original position is the means for answering the question of what justice requires, for Geuss it is the original position itself that needs to be scrutinised, asking why a particular society, at a particular time, should come to hold that the way to think about justice is to imagine away everyone's actual position in the world. Thus, the objection cannot be answered by showing that Rawls's theory works on its own terms. Geuss's point is that we need to justify the decision to make this abstraction in the first place. The defence offered above, that the original position transcends the status quo precisely by excluding endowments, is thus for Geuss a restatement of the charge rather than an answer to it. This objection, however, belongs to political realism proper, and is taken up in chapter 3.

2.7 TO CONCLUDE: DOES THE POLITICAL HAVE TO BEGIN WITH MORALITY?

In this chapter, I have examined Rawls's political liberalism, with particular attention to its relationship with morality. Let us now briefly summarise why and how morality serves as a foundation for Rawls's approach. Furthermore, I shall defend the claim that, contrary to the critique advanced by political realists, Rawls's justice-driven politics is not hopelessly disconnected from the practical realities of politics. Rather, it is essential for the kind of society he envisions politics ought to strive for – namely, “stability for the right reasons” (1996, 391), as opposed to “only a *modus vivendi*, and society's stability depend[ing] on a balance of forces in contingent and possibly fluctuating circumstances” (ibid., 392, original emphasis).

As James Gledhill observes, Rawls's ideal normative theory does not ask us to imagine a utopian world; rather, it urges us to recognise that “what separates us from realizing a more perfect society is forging the collective practical political will to do so” (2012, 82). Rawls does not aim to uncover the substantive content of what constitutes a ‘more perfect’ society; instead, he emphasises that the existence of multiple comprehensive doctrines is not a problem to be solved, but

a condition of pluralism that every reasonable person must learn to navigate. Crucially, these doctrines are understood to lie beyond the political sphere. This distinction between the political and the personal is fundamental to Rawls's political liberalism. When agreement is sought, it is not a matter of one worldview dominating the discourse, as realists fear (Geuss 2008), but rather of reasonable citizens exercising their moral and epistemic capacities to determine which elements of their comprehensive doctrines may contribute to a conception of justice acceptable to all. Such is the nature of justice – justice grounded in the *right reasons* – that underpins the kind of social stability towards which the political process is directed, as we shall explore next.

2.7.1 Justice as the source of stability for the right reasons

Stability is a central concern in Rawls's political philosophy. His political framework seeks to ensure stability by cultivating both the internal "sense of justice" in individuals (1996, 103–104) and their need for social cooperation. Rawlsian stability goes beyond mere acceptability or legitimacy; it aspires to a deeper and more enduring form of stability. This involves not only a degree of security and predictability in political arrangements, but also *genuine* support from within the citizenry.

Such stability does not emerge simply because it is desirable as a prerequisite for political action (though this is also true); rather, it arises from collective action conducted in accordance with principles of justice. It emerges from within the framework of justice and must be achieved through shared interaction that recognises all individuals as moral equals. Stability is not something imposed from outside; it is created by citizens themselves.

Put differently, Rawls envisions a society that is stable for *intrinsic* reasons – reasons that stem from justice itself. Justice is not external to society, nor is it confined to specific institutions or groups. To paraphrase Foucault's (1978, 93) well-known claim about power, justice is everywhere, not because it embraces everything, but because it comes from everywhere. It arises from the moral and epistemic capacities of citizens (see 2.5.2) and from the idea of moral truth as 'unowned' (2.4.2). For Rawls, political power, despite always being coercive, derives from the moral and rational capacities of individuals.

In this framework, Rawlsian justice follows a particular sequence, beginning with individuals, moving towards social consensus, and culminating in political stability. Each stage presupposes the one before it, so that there is a capacity for justice in persons, a consensus grounded in that capacity, and a stability that arises from such consensus. Crucially, the consensus Rawls envisions rests on a

relational conception of justice, one that emerges not from individuals in isolation but from their interaction and mutual recognition as equals.

Two central questions follow from this. First, if a just society is stable for the right reasons – i.e., because citizens genuinely support it – what, then, are the wrong ways to achieve stability? To answer this, we must revisit the concept of *modus vivendi*, briefly introduced earlier. Rawls contrasts his view of overlapping consensus with *modus vivendi*, a form of political agreement based on “accepting certain authorities, or on complying with certain institutional arrangements, founded on a convergence of self- or group interests” (1996, 147). In such arrangements, stability is merely a contingent equilibrium, since no party has an incentive to disrupt it under current conditions, but any party may pursue its own interests at others’ expense once circumstances change. This kind of stability is inherently fragile, as it is grounded solely in mutual advantage.

A political order in which parties act only on the basis of their own interests – disregarding “how these affect social concord and the moral quality of public life” (1996, 146–147) – lacks the moral foundations needed for lasting stability. By contrast, a political order grounded in an overlapping consensus on justice is stable even amid shifting power dynamics (ibid., 148). Within such a consensus, greater power does not justify unilateral renegotiation or imposition of a majority view. For Rawls, stability based on justice is genuine and self-sustaining. It does not depend on force or manipulation but is upheld by public transparency and the willingness of citizens to subject political principles to open critique (see also 2.5.2). As Rawls puts it, in “public political life, nothing need be hidden [...] there is no need for the illusions and delusions of ideology for society to work properly and for citizens to accept it willingly” (ibid., 68–69).

The second question concerns what constitutes the right reasons for stability, and how these reasons might be realised. To answer this, we return to Rawls’s conception of the political actor (2.5). The motivation to reach an overlapping consensus on the political conception of justice is grounded in the moral and epistemic powers of individuals. These include the capacity for a conception of the good and a sense of justice – in short, a “moral personality” defined by both rationality and reasonableness (1996, 302). Importantly, this moral personality differs from a merely calculative or strategic form of rationality.

The rational aspect of this personality refers to one’s capacity to develop a “conception of the good” as “a system of final ends” (1996, 280) – that is, an understanding of what constitutes a worthwhile life (ibid., 302). Rawls assumes that every rational agent possesses such a conception and therefore cannot be devoid of moral capacity. All individuals require a value framework to make decisions that are not only materially beneficial but also meaningful in relation

to a sense of purpose and membership in a shared moral order. Someone who lacks a conception of the good cannot be considered fully rational and is therefore unfit to participate in the process of consensus-building as a moral equal (ibid., 375). Notably, Rawls's notion of the 'good' is not reducible to desires understood as self-interest or instrumental utility. In this sense, his conception of human nature diverges fundamentally from Hobbes's, who grounds political order in power and self-interest (Grcic 2007, 390).

The second moral power – reasonableness – is defined as the “capacity to honour fair terms of cooperation” (1996, 302), particularly in the context of pluralism and disagreement. Reasonableness is relational; it involves engaging with others on the basis of reciprocity and recognising them as moral equals. A reasonable person seeks mutual solutions rather than imposing their own view, and is “prepared to offer one another fair terms of social cooperation ... and to agree to act on those terms, even at the cost of their own interests in particular situations, provided that others also accept those terms” (ibid., xlv). To possess a moral personality, then, is to be capable of considering others' perspectives and distinguishing between personal commitments and those that must be translated into political conceptions insofar as they relate to shared cooperation.

The “burdens of judgment” (1996, 54–58) further cultivate this moral sensibility by fostering *epistemic respect* – i.e., respect for others' capacity to reason and form rational beliefs, rather than the substance of their beliefs and commitments. This respect entails recognising the legitimacy of others' normative commitments and granting them due consideration. Such epistemic respect motivates cooperation from within comprehensive doctrines themselves; it is not introduced externally by the political system. Rawls argues that all reasonable doctrines share this feature.

Rawls's political liberalism is a contractarian ideal theory that seeks to establish principles to which all reasonable citizens could hypothetically consent. The possibility of such agreement depends on the assumption that it is grounded in something beyond mere coercion. Where coercion is absent, participation must be voluntary; and voluntariness, in turn, requires a willingness to accept shared consequences and assume responsibility for their outcomes. As Leader Maynard (2025, 15) suggests, morality is inherently “other-regarding”. Without an obligation to recognise such responsibility towards others, the basis for sustaining the agreement would be reduced to self-interest, that is, acting morally only for one's own sake.

In this sense, the civic obligation presupposed by the consensual agreement Rawls envisions has a moral foundation. Without this moral dimension, concern for others would be absent, and agreement on shared principles would reflect not

genuine voluntariness but merely self-interested and instrumental reasoning. However, this moral dimension is not substantive but conceptual. It does not depend on agreement about a particular moral doctrine; rather, morality is embedded in Rawls's conception of persons as ends in themselves and as possessing the two moral powers (see 2.5.1), as well as in the conditions of fair cooperation, including the burdens of judgement and the original position.

I have argued further that each comprehensive doctrine requires a degree of opposition from others to maintain its coherence and relevance. This coherence arises not in isolation but through the tensions generated by engagement with alternative worldviews. Accordingly, Rawls's idea of overlapping consensus does not seek to eliminate doctrinal diversity, but rather to sustain it within a *shared political framework*. It is the ongoing pursuit of consensus – not the attainment of full agreement – that carries political significance (see Gledhill 2012; Thomas 2017). Every belief or commitment acquires coherence through contrast; thus, a minimal level of cooperation becomes one of the rational goods that Rawlsian political actors are motivated to preserve.

For this reason, I have claimed that once justice enters the political domain, its 'truth' becomes 'unowned' (2.4.2). Justice does not arise *ex nihilo* from empirically observable commitments, as Hobbes (Grcic 2007, 390) or political realists like Geuss (2008) argue. Nor is it a neutral or empty signifier. Rather, the conception of justice – but not its content – becomes a shared standard that reasonable persons could accept. While the justification for shared political justice draws on individuals' comprehensive moral doctrines, these doctrines provide structural, not substantive, content to the shared political framework.

To conclude, stability for the right reasons rests not on shifting alliances or mutual interest, as in a *modus vivendi*, but on a shared conception of justice that citizens endorse through their own moral capacities. This conception of justice endows citizens with a self-ascribed responsibility to engage in ongoing consensus-seeking. In light of the persistent diversity of comprehensive doctrines, which Rawls considers not a problem but a desirable, and indeed inevitable, feature of democratic life (1996, xviii), this process is necessarily unending. Rawls's ideal theory, then, does not aspire towards a definitive moral vision awaiting realisation; its idealisation is heuristic rather than telic (2.6.1), supplying citizens with a register within which to compose their own conception of justice, not a destination political practice is meant to approximate. Stability for the right reasons is accordingly never a state achieved once and settled, but a condition sustained only for as long as citizens continue to affirm it from within their own comprehensive doctrines.

2.7.2 Lessons from Rawls

Rawls's theory seeks to justify a particular form of political arrangement, one constrained by a moral principle of justice. A just government is more likely to garner broad support and legitimacy, thereby ensuring societal stability. This is the instrumental justification for Rawlsian political liberalism, since justice serves as the foundation for political arrangements because it secures both stability and legitimacy.

Yet, there is also a deeper justification for viewing a just state as more than an aspirational ideal. This lies in Rawls's assumption that individuals are not merely self-interested agents striving to maximise personal gain regardless of its impact on others. Rather, an intrinsic sense of justice motivates us to resolve disagreements while remaining true to our diverse foundational values. The process of working together to agree on how to live collectively reflects our profound moral commitments and affirms the idea that moral norms gain meaning within the context of human relationships – whether among individuals, social roles, or institutions.

Building further on Rawls, I argue that in the context of persistent value disagreement, a shared moral grounding enables a form of reasonable cooperation, one that *neither eliminates nor reconciles* divergent moral systems, but builds upon them. On this view, the very plurality of comprehensive moral doctrines is what makes genuine coexistence possible, not mere survival. We should therefore resist suppressing or homogenising our moral commitments, since these commitments sustain the diverse identities essential to both personal and collective life. Political justice, then, should be understood as a unifying framework that enables common life amid difference.

At first glance, this may appear to be an overly romanticised interpretation of Rawls's theory, based on idealistic assumptions about the functioning of modern societies. Yet this is not the case. I depart from interpretations that frame Rawls as concerned primarily with ideal moral ends – what Matthew describes as Rawls's effort to determine “what a (nearly) perfectly just – yet achievable – society looks like” (2019, 556). In my reading, Rawls's idealisation is heuristic rather than telic (2.6.1), as he does not assume that societies are or will become just, but only that citizens share a sense of justice strong enough to keep them seeking fair terms.

This chapter has shown how Rawls's reliance on a pre-political but non-foundationalist morality makes his political liberalism compelling. Removing moral principles from Rawls's theory would undermine what, in his view, makes living together possible. Without a shared moral basis, we would lack the intrinsic

motivation to create arrangements acceptable to all. Instead, we would rely solely on interest-driven compromises, which fail to sustain core aspects of our humanity, particularly respect for moral equality and freedom.

Instrumental reasoning alone, being externally motivated, is insufficient to inspire individuals to view one another as free and equal. For someone to be recognised as free, they must be acknowledged as capable of choosing their own fundamental commitments – provided these are reasonable and mutually acceptable. From the recognition of moral equality, other forms of equality follow. For example, when a prison guard and a prisoner regard each other as moral equals, they can accept the terms that govern their respective roles – so long as those terms respect their shared moral status. This respect affirms their dignity, grounded in Rawlsian notions of “moral personality” ([1971] 1999, 442). Conversely, when power relations – whether between prisoners and guards or states and citizens – lack a shared moral foundation rooted in mutual respect, power is more likely to be abused in the pursuit of self-interest. From a Rawlsian perspective, without pre-political morality, politics risks becoming a game of chess, where the most strategic player wins and others lose. And if we accept this view, we can also accept Rawls’s conception of society as a non-zero-sum game; it should aim to benefit everyone, not merely the victors.

What Rawls offers, then, is not a prescription for the content of justice, and in this respect he is more restrained than his realist critics allow. But the moral conditions under which citizens may determine that content – the original position, the veil, and the two moral powers – are still settled before any citizen determines anything, and this is where, conceptually speaking, his priority of value over valuation lies.

3. POLITICAL REALISM

3.1 WHAT IS POLITICAL REALISM?

This chapter examines contemporary political realism as the first significant challenge to political moralism, while the second challenger – agonism – will be explored in chapter 4. Contemporary political realism (hereafter referred to simply as political realism) represents a revival of classical realist thought, associated with figures such as Machiavelli, Hobbes, Hume, Weber, and Schmitt (McQueen 2017). This resurgence – most notably developed by Bernard Williams (2005) and Raymond Geuss (2008) – defines itself in opposition to the Kantian-influenced traditions said to dominate contemporary political philosophy.

Realists critique these approaches by rejecting the view that normative theorising about political arrangements should be grounded in pre-political moral ideals. Their critique is directed in particular at the political philosophy of modern Rawlsian liberalism (Leader Maynard 2025, 13), which has often been regarded as the mainstream of political theory, or what William Galston (2010) refers to as “high liberalism”⁸. However, this critique may also apply to other normative approaches, such as communitarianism, when they treat moral values as existing before politics rather than as shaped by historically and politically situated concerns.

Realists reject political theory that prioritises moral norms as antecedent to politics because they regard such norms as products of social and historical contingency rather than as reflective of objective value positions. For realists, this renders such theorising dogmatic, its most significant flaw being the concealment of this contingency under the guise of universality. As Sleat writes, “what is wrong with political moralism is not that it represents a form of political theory that starts outside of politics, but that it thinks it does” (2022, 473, original emphasis). Similarly, [Kreutz \(2023, 14–15\)](#) notes that realists consider morality to be particularly prone to infestation by political power, shaped in the image of the latest political victor.

While realists remain vague in their characterisation of ‘moralism’, they target theorists who, in various ways, regard justice as the primary aim of politics (Leader Maynard 2025). It will suffice, for present purposes, not to delve into the specificities of the different strands of moralism, but rather to adopt Leader Maynard’s succinct definition, whereby “practical normative judgements about

⁸ Other notable scholars within this tradition include Ronald Dworkin (2011), Martha Nussbaum (2011), and T. M. Scanlon (2000).

politics – about what political actions or norms or institutions are good, right, or permissible – depend, in some sense, on moral judgements” (ibid., 14). As such, political realists are perhaps best understood as ‘anti-moralists’.

Rather than conceiving politics as a domain for addressing questions of citizens’ well-being, justice, or the good life, realists maintain that political theorising should concentrate on strictly political arrangements – such as configurations of power relations and collective decision-making – eschewing excessive or inherently contestable moral commitments (Rossi and Sleat 2014). For realists, the notion of prioritising justice before politics is incoherent, as they argue there exists no objective foundation for determining a universally ‘correct’ or ‘objective’ conception of justice (Hall and Sleat 2017). Accordingly, instead of subjecting political practices to external ethical standards, realists advocate an emphasis on legitimacy, i.e., the exercise of authority constrained by political procedures (Williams 2005; Cross 2023), contending that political relationships should neither be subsumed under moral aims nor reduced to mere instruments of domination (Rossi and Sleat 2014).

Political realism therefore regards politics as a distinct domain of social life, governed by its own principles and normative standards. It is sufficiently wide-ranging that it might be more appropriately characterised as an approach to theorising which, above all, prioritises a true-to-life perspective on politics. Alison McQueen explains this more comprehensively in her formulation, according to which

“[political realism is a] family of approaches to the study, practice, and normative evaluation of politics that (a) affirms the autonomy (or, more minimally, the distinctiveness) of politics; (b) takes disagreement, conflict, and power to be ineradicable and constitutive features of politics; (c) rejects as ‘utopian’ or ‘moralist’ those approaches, practices, and evaluations which seem to deny these facts; and (d) prioritizes political order and stability over justice (or, more minimally, rejects the absolute priority of justice over other political values)” (2017, 297)

Let us now examine this in detail. First, political realism is (a) not a single theory, but rather a family of approaches that broadly resist the reduction of normative thought to non-political foundations, most notably moral philosophy. Although this family encompasses diverse definitions of the political, we can distinguish between *radical* and *moderate* strands concerning political theory’s relationship to ethics and moral norms. Radical realism rejects the primacy of moral norms over politics, both external to and within the political domain, whereas moderate realism conceives of politics as distinct from, though not entirely divorced from, the realm of values. While both strands are committed to contextualism and tend

to put political order before justice (McQueen 2017, 302–306), they differ in how they engage with ethics. Radical realism rejects any ethical foundation for politics, on the grounds that such foundations distort political reality, whereas moderate realism allows ethics a place within the political sphere, though not as a primary foundation.

Most realist theorising aligns with the moderate strand, typically following in the footsteps of Bernard Williams (2005). Radical realism, initially advanced by Raymond Geuss (2008), has gained comparatively limited traction (Sleat 2022, 472) and has been subject to substantial criticism for its selective reasoning (Baderin 2021; Erman and Möller 2015, 2022, 2023b, 2023c, 2025; Leader Maynard and Worsnip 2018; Leader Maynard 2025). Furthermore, contemporary literature within the radical tradition has thus far primarily focused on a specific form of political theorising – namely, ideology critique (Prinz and Rossi 2017; Aytac and Rossi 2023) – whereas the moderate approach, in recognising the complex interrelation between the political and ethical domains, offers a more extensive body of work for critical engagement. For example, Baderin (2021) argues that realism’s separation of politics from morality obscures morally relevant and actually existing considerations, such as epistemic injustice. In a similar vein, Elizabeth Frazer (2018), in her discussion of feminism and political realism, observes that feminism, being inherently realist, ought to be grounded in what she terms philosophical realism.

Second (see points b–d in the quotation above), McQueen explains that the objective of realist inquiry is to investigate how disagreement, conflict, and power shape or distort political order and stability – rather than treating politics as a vehicle for realising justice or other ethico-political ideals. For realists, politics is not primarily concerned with striving for a better life – an aspiration they regard as abstract and utopian, but, perhaps more importantly, also dogmatic (Geuss 2008) – but rather with aligning our understanding of collective life with its actual conditions. To represent political reality adequately, realists argue that politics should be examined as an autonomous domain, though not one entirely isolated from other spheres of life. Accordingly, when addressing the normative question of how politics might be improved, realists emphasise organising concepts – such as power, authority, autonomy, ideology, consent, legitimacy, and participation – over substantive ideals like democracy, citizenship, freedom, and equality, which they view as susceptible to doctrinal or ideological distortion. In other words, realist normative theorising allows engagement with moral considerations only after attending to what Bernard Williams described as the “first political question”: “securing order, protection, safety, trust, and the conditions of cooperation” (2005, 3; see also Galston 2010, 388).

After this brief introduction to political realism, I now turn to its distinctive features. I begin by examining the rationale behind the realist critique of moralism (3.2). I then consider the difficulties realists face in articulating normative commitments while rejecting pre-political values as the primary standard for evaluating political arrangements (3.3). I conclude by arguing that this kind of realist approach risks reducing politics to an end in itself (3.4).

3.2 REALISM AS ANTI-MORALISM

Political realism is the most sustained attempt yet made to dislodge morality from its position of priority in political thought (Williams 2005; Geuss 2008; Rossi and Sleat 2014). If the realist critique succeeds, my argument in chapter 6 – that politics without pre-political values lacks the resources to envision and imagine a different and better society – would be largely redundant. The solution would simply be to abandon pre-political values and treat politics as a tool for pursuing political ends alone. But I argue that the realist critique of politics as “applied ethics”, as Geuss (2008, 6) puts it, does not succeed, even on its own terms.

The realist critique of moralism does not constitute a rejection of morality *per se*, but rather a rejection of the claim that morality should be situated prior to and above politics. Edward Hall describes political realism as “grounded in a rejection of moral theory as standardly conceived”, emerging “from the adoption of a realistic spirit in ethics that problematises traditional accounts of morality, as well as the assumed view of the relationship between morality and politics that such accounts explicitly or implicitly endorse” (2020, 3). On Hall’s account, the source of the moralists’ error lies in “a profound misunderstanding of ethics” (2020, 2). For realists, political theorising should remain within its own distinct boundaries – those encompassing institutions, legitimacy, and the “distinctive goods of political life” (*ibid.*).

It is therefore worth examining what, exactly, realists take this misunderstanding of ethics to be, i.e., what is the traditional account of morality that a “realistic spirit in ethics” (Hall 2020, 3) seeks to problematise? Hall and Sleat argue that the impulse behind both Williams’s and Geuss’s realism was an attempt “to philosophise about politics without relying on understandings of morality which we have little reason to endorse” (2017, 280). Realism, on this account, is not anti-evaluative *per se*; it objects not to the presence of values in political thought, but to placing moral ideals before politics.

Realists are not necessarily short of positive ethical resources. Williams (1985), representing the moderate strand, offers thick ethical concepts, the ethical federation, and a conception of the respective demands of personal and public

morality. In particular, he characterises personal morality as relying on thick concepts and public morality as relying on thin ones, since thin concepts are more abstract and so better suited to wide public justification, whereas thick concepts are richer in content and better capture particular, personal ethical experience (2005, 48–50). For Williams, Kantians and utilitarians attempt to unify these two spheres under a single system by prioritising thin justification, but he contends that this merger fails because it overlooks the different requirements of personal ethics, which call for rich, context-sensitive concepts, and public morality, which, in diverse societies, must rely on abstract, procedural norms of legitimacy in order to function at all. Thus, Williams proposes that we should not try to apply a single moral framework to every area of life. Instead, we should have a more pluralistic ethical structure – an “ethical federation” (ibid., 50) – in which thicker moral ideas guide particular communities, while thinner, procedural principles guide public discussion.

What moderate realists struggle with, however, is integrating these resources into political theory, while the radical strand does not attempt to develop a positive political theory at all, as Geuss readily acknowledges (2008, 95–96).

Having outlined the realist dissatisfaction with a political philosophy that treats values as the evaluative standard of politics, we can now turn to the realist conception of morality, beginning with the moderate strand (3.2.1) and then the radical strand. The radical strand is best approached through ideology critique (see 3.2.2), which, in what I take to be a productive move, treats morality as the means by which power justifies itself, in a way that continues the genealogical critique this study endorses (see 1.3). The section then turns to the realist critique of Rawls (3.2.3), arguing that the moralism it attacks is not the one Rawls actually holds.

3.2.1 Williams: morality as a “peculiar institution”

While Nietzsche’s critical genealogy exposed the historical sources of modern morality (1.3), Bernard Williams terms morality a “peculiar institution” (1985, 174–196), which operates under the pretence of objectivity, obscures the complexity of ethical life, and imposes an artificial hierarchy wherein morality is regarded as superior to all other forms of reasoning, including the political. It is peculiar because, for Williams, it is internally contradictory, imposing demands that are simultaneously moral in form and, in practice, potentially immoral. He says:

“Many philosophical mistakes are woven into morality. It misunderstands obligations, not seeing how they form just one type of ethical consideration. It misunderstands practical necessity, thinking it

peculiar to the ethical. It misunderstands ethical practical necessity, thinking it peculiar to obligations. Beyond all this, morality makes people think that, without its very special obligation, there is only inclination; without its utter voluntariness, there is only force; without its ultimately pure justice, there is no justice. Its philosophical errors are only the most abstract expressions of a deeply rooted and still powerful misconception of life” (1985, 196).

Williams argues, then, that the morality system is oriented towards obligation, in that moral deliberation aims at a reasoned conclusion that, once reached, must be acted upon. The resulting sense of obligation rests on the principle that ought implies can, given that “If my deliberation issues in something I cannot do, then I must deliberate again” (Williams 1985, 175). On this basis, morality is constrained by the moral agent’s capacity to enact the decision.

But more importantly, for moral decisions to be actionable, they cannot be mutually conflicting (1985, 176); that is, there can be only one truly moral course of action, otherwise moral loss and consequent blame arise (ibid., 177). As Williams observes, “Morality encourages the idea, only an obligation can beat an obligation” (ibid., 180, original emphasis). The function of obligation, according to Williams, is to “secure reliability” (ibid., 187), since making a promise or establishing a principle is meant to guarantee a future course of action, so that when A happens, B follows. This mechanism provides certainty and determinacy in social relations, ensuring that individuals share an understanding of their common future.

However, it does not follow that obligation and trust – both vital to social interaction – are intrinsically connected. This is so for two reasons. First, trust is unnecessary if moral obligations are invariably fulfilled, since trust presupposes an element of uncertainty. Second, within that space of uncertainty, trust may indeed benefit from mutual responsibility, but only when all parties share the normative framework that prescribes the relevant obligation.

Furthermore, the ought-implies-can principle entails that moral decisions must be voluntary; morality cannot, therefore, be imposed. As Williams writes, “The moral law is more exigent than the law of an actual liberal republic, because it allows no emigration, but it is unequivocally just in its ideas of responsibility” (1985, 178).

There is, however, a further twist in Williams’s argument. Beyond the deadlock within deontological rationality created by its emphasis on obligation, Williams

identifies a similar impasse in utilitarianism. He argues that by prioritising overall well-being above all else, utilitarianism renders moral considerations merely one set among many competing evaluative criteria. Consequently, morality cannot be understood as obligation, as some utilitarians claim, since for a utilitarian end to be viable, all values must be regarded as instrumental rather than intrinsic (Williams 1985, 194), thereby ultimately calling into question the very value of moral reasoning within the utilitarian framework itself.

What renders the morality system incoherent, on Williams's account, is the intersection of universality, obligation, and voluntariness, which together commit it to the ambition of transcending luck (*ibid.*, 195), and so to constructing fictional worlds capable of sustaining its authority (*ibid.*, 196). To be applicable at all, morality must tacitly supply the assumptions and select the features of a situation under which its principles appear to apply. Its claim to purity, Williams argues, is secured by abstracting moral consciousness from emotional response and social influence, and this abstraction conceals both how the system disciplines those who deviate from it and whatever merit those disciplinary means might possess. Such concealment is not accidental, as these merits could be recognised only from outside the system, from a perspective that could judge its value. But the system is closed in on itself and treats the use of any values other than its own as, in Williams's words, an "indecent misunderstanding" (*ibid.*, 195).

Morality thus becomes an end in itself, arranging its circumstances to suit its methods. Because the method relies on rational deliberation, it separates moral judgement from anything that might interfere with that process, such as emotion, social norms, or personal relationships. By doing this, it makes it difficult to criticise the system from within, and underlying mechanisms can only be identified from an outside perspective. The result is a morality that is self-justifying (see also [Aytac 2022](#)), and that can register any challenge from beyond its own terms only as a category mistake.

Williams further argues that morality is not only an illusion in itself, but also a system that creates further illusions, drawing us further away from the idea that social justice, which morality purports to tackle, always contains an element of luck. Yet Williams also argues that luck, which the moral system claims to correct (with its purity of definitive answers), cannot be entirely independent of values. However, this does not mean that Williams calls for abandoning justice altogether, as without justice, we would be left with only utilitarian efficiency, domination, or chance. Instead, he calls for rethinking moral and political values in a world where luck is always present, without falling into cynicism or a politics based simply on power (*ibid.*, 196).

3.2.2 Realist ideology critique

For realists of the radical strand, morality is especially vulnerable to capture by political power, since once politicised it becomes ideological and is reshaped in the image of whoever holds power at the time (Kreutz 2023, 14–15). Such distortion takes one of two forms, either wishful thinking, which shapes belief to fit one's own interests, or ideology in the strict sense, which presents those interests as objective (Geuss 2017, 241–242, 245). The task of realist theorising, accordingly, is to expose how moralisation manufactures and entrenches the promotion of particular interests (Prinz and Rossi 2017; Aytac and Rossi 2023; Clark 2024).

This idea draws on the same conceptualisation of morality employed in this study (see 1.4), which is rooted primarily in Nietzsche's genealogical account ([1887] 2007; see also 1.3). The overlap is unsurprising, since Geuss (2008) likewise draws on Nietzsche in arguing that power justifies itself through epistemic distortion by constructing stories of legitimation (Rossi 2025). Morality is particularly well suited to this purpose because, when treated as intrinsically valuable and requiring no further justification, it is considered a foundational – i.e., unquestioned – source of legitimacy. To illustrate this, consider Geuss's comment on liberalism: “one way ... in which a [liberal] political philosophy can be ideological is by presenting a relatively marginal issue [e.g., of justice] as if it were central and essential” (2008, 54). Thus, on this view, what ethics-first theorists take to be the purpose of politics is the instrumentalisation of moral power, where morality is constructed to function as the glue holding existing structures together. The aim of realist ideology critique is to uncover these distortions by examining how political systems justify themselves by treating circular reasoning as a sign of ideological justification, regardless of which particular values are being used.

Realists argue that, in order to identify how ideologies misuse power, we must employ immanent critique. Immanent critique, as Stahl defines it, is “a form of social critique that evaluates both the empirical behaviour constituting social practices and the explicit self-understanding of their members according to standards that are, in some sense, internal to those practices themselves” (2022a, 7), aiming at a transformation that encompasses both action and self-understanding. It differs from external critique, which brings in norms from outside the system and risks dogmatism, and from internal critique, which only checks whether a system's norms are consistent with its practices without questioning either of them (Stahl 2013, 8). Immanent critique aims at something more ambitious than either, namely to identify from within a system the distortions that allow it to sustain itself, while explicitly setting its own moral commitments aside (Prinz and Rossi 2017; Stahl 2017), treating morality itself as

one of the objects of critique – a potential source of self-justifying power – rather than a standard the critique relies on (Stahl 2022a; 2022b, 13). Applied to ideology, it identifies the sources of injustice by observing normative commitments that are direct products of coercive power within the political system, examining how power structures manifest in societal norms, values and practices in order to reveal how these perpetuate inequalities and injustices (Prinz and Rossi 2017; Aytac and Rossi 2023).

Aytac and Rossi (2023) argue that the problem with ideology is epistemic, not moral, since dominant groups spread self-serving beliefs about their own legitimacy among subordinate groups, which produces a circular and epistemically flawed picture of social relations (*ibid.*, 1215). On this view, the problem with ideology is not that it sustains hierarchy as such, but that it sustains hierarchy through motivated reasoning; and motivated reasoning, not moral wrongness, is what should distinguish malign cases of self-justification from benign ones (*ibid.*, 1219, 1221). Crucially, they explicitly reject lived experience and moral intuition as sources of evidence, on the grounds that these already assume the kind of objective moral standards realism denies exist (*ibid.*, 1217–1218). In their view, incorporating such assumptions risks introducing new ideologies, potentially perpetuating inequality. Radical realists therefore reject socialist or feminist ideology critique for obscuring reality by substituting one social construct for another (*ibid.*, 1218). Instead, their method is to analyse empirically how power constructs and disseminates self-legitimising narratives (*ibid.*, 1219–1220), offering persistent scrutiny of power without a predetermined outcome.

Having outlined the aim and scope of radical realist ideology critique, we should also highlight its costs, which stem from a pursuit of value neutrality that Aytac and Rossi do not directly address (but see 2023, 1220–1221, 1224–1225). Specifically, I argue that radical realists reduce the scope of politics to the bare minimum necessary for a society stable enough to avoid conflict and possessed of basic mechanisms for maintaining public support (*ibid.*, 1224). In their case study demonstrating epistemic distortions in neopatriarchal ideology, Aytac and Rossi assert that “it is the epistemic distortion generated by this motivated reasoning that we deem problematic, not the existence of privilege and hierarchy per se. This is not a moral condemnation... but a warning that powerful elites tend to overestimate the functional necessity of hierarchy” (*ibid.*, 1225). By aiming for a political order that merely ensures society’s basic stability, however, we deprive ourselves of the opportunity to envisage progress. If the criteria of epistemic critique for detecting motivated reasoning and circular argumentation are measured against this minimal standard, we are left with an impoverished understanding of what constitutes a desirable society. There is, furthermore, a contradiction in terms, whereby the critique is intended to be disruptive yet lacks

transformative ambition, taking political stability as its point of reference. Distortion, in short, can be detected without a standard of value; improvement, I argue, cannot be envisaged without one.

The pursuit of value neutrality implies that radical realists envisage a politics devoid of ideology, as Sleat also notes (2024, 155–156). Yet as Freeden emphasises, politics would lose its meaning were it not thought of in terms of ideologies – not necessarily pejoratively, but as the decontestation, or fixing, of the meaning of concepts so as to create coherent constellations of worldviews in politics (1996, 4). Politics is thus fundamentally concerned with the diversity and interplay of ideologies, rather than being a mere aggregation of unordered preferences. What realists perceive as a distortion of reality is in fact an essential aspect of ideological engagement, since the process of decontestation centres precisely on differing conceptions of what ought to be considered reality. Politics is not about exchanging make-believe stories, but about navigating a plurality of understandings of what is true.

The issue extends beyond the contestation of the substance of truth to the epistemics of acquiring it. The notion of a ‘politically innocent’ epistemology of truth is untenable, intertwined as it is with historical struggles for power, as Sleat rightly notes, further adding that the aspiration to an ideology-free political reality is itself ideological (2024, 156). A purely epistemic approach to critiquing ideology may not, then, be devoid of the distortions engendered by power, because power is not solely a matter of self-justification and political malice; it is also self-deception. If the critic analysing epistemic strategies is not fully transparent about whose expertise they draw upon and the sources thereof, and is not reflexive regarding their own thought process, a suspicion will always linger regarding the ‘purity’ of a ‘purely epistemic’ critique. Since no critique can be completely severed from the object of its criticism, ignoring the substantive content of ideologies also undermines the reflexivity necessary for epistemic work. Rather than aiming for epistemic purity, then, realist ideology critique would perhaps be more effectively conducted through the prism of epistemic integrity – a standard which asks of the critic what the critique asks of its objects, and which is, notably, a virtue rather than a rule.

The attempt to exclude value premises prior to critique, and the impossibility of doing so, have been heavily criticised (Baderin 2021; Erman and Möller 2015, 2022; Leader Maynard and Worsnip 2018; Sleat 2024, 139). At minimum, ethical appraisal inevitably – albeit implicitly – constitutes part of the process of selecting cases for critique. Realist ideology critics (Prinz and Rossi 2017; Rossi and Argenton 2021; Aytac and Rossi 2023) focus, after all, on cases where an ideology has already been identified as ethically problematic by scholars from adjacent disciplines. The flaw therefore lies in the fact that, although Aytac and Rossi by

their own admission do not advocate a one-size-fits-all approach to ideology critique, they nevertheless explicitly aim to present a non-moral version of it and ultimately fail to do so. Given that radical realist ideology critique employs the method of immanent critique (Prinz and Rossi 2017) and shares its emancipatory aims, its motivations must be externally grounded (see also Stahl 2022b); otherwise the critique remains constrained by self-imposed standards, since the substantive aim of emancipation is inherently moral in nature (Meijer 2022; also Allen 2001; but for arguments for non-moral emancipation, see Buchanan and Powell 2017). The same conclusion is reached from the opposite direction by contemporary critics of ideology, who are open about their normative (moral) commitments. As Chambers observes, to identify ideological oppression one must set out the evaluative principles by which an ideology is judged oppressive, and this requires a “strong and unashamed commitment to substantive, controversial normative values” (2017, 177–178).

Others have identified further flaws in realist ideology critique, which I regard as contributing to its substantive failure to deliver the emancipatory or progressive critique it purports to offer. Sleat (2024, 145) and Clark (2024, 263–264) observe that determining whether *malignant instances* of circular justification are indeed politically motivated, as Aytac and Rossi (2023) suggest, is problematic. They demonstrate that it is not possible to directly infer from an act of indoctrination by a dominant group that the underlying motivational reason is political power, as such actions could just as easily be driven by other factors (Sleat 2024, 146–147). The inference rests on the assumption that political actors in dominant positions are consistently rational and self-interested agents, making conscious and calculated choices among various alternatives, and it neglects the possibility that these actions may also be rooted in fervently held religious beliefs.

As Clark (2024, 266–268) also observes, it is insufficient to focus exclusively on justificatory narratives that appear circular; attention must likewise be directed towards *alternative justificatory frameworks* that contribute to ideological distortion. Put simply, radical realist ideology critique is overly rigid, thereby undermining its plausibility – an outcome that realism, after all, seeks to avoid. Moreover, the critique is ineffective because it is rigid, and it is rigid because it insists on value neutrality and takes political stability as the aim. To address this limitation, it is necessary to subject all justificatory narratives to critical scrutiny, not solely those that uphold hegemonic structures, in order to prevent politically motivated reasoning from emerging within alternative forms of circular argumentation.

Radical realist ideology critique therefore fails to be genuinely radical – that is, truly transformational – because it directs its focus towards a false target. It

remains preoccupied with avoiding moral premises at all costs, thereby neglecting the fact that moral ideals, or the moralisation of politics, constitute only one dimension of the operation of power. In this respect, it falls short in its critique of power, as it concentrates its efforts on attacking moral norms which, in Geussian terms, represent “dead politics” (Geuss 2008, 42) – norms that have been solidified and instrumentalised to sustain dominance over time – while failing adequately to address other ideational distortions rooted in politically motivated economic or cultural justifications of power. In effect, radical realist ideology critique politicises *morality*, reducing it to a mere instrument that corrupts what it construes as ‘pure politics’. As a result, it capitulates to the very orthodoxies it aims to interrogate and transcend. By adopting social stability as the standard of political equilibrium, radical realist ideology critique forgoes more ambitious, yet still realist-compatible, conceptions of society. Instead, radical realists become ensnared in a Hobbesian-Geussian pessimism that resists a progress-oriented understanding of politics (Honig and Stears 2011).

None of this, however, takes away from what realist ideology critique does well, namely, provide a convincing account of how power uses moral language to justify itself. Any theory that gives values a role in politics must also explain how those values can be protected from being manipulated by those in power, and this also requires intellectual honesty and humility, as critics must be willing to subject their own views to the same scrutiny they apply to others. Chapter 6 takes up this idea as a virtue of political judgement. In short, realist ideology critique locates morality in the right place, among the means by which power can justify itself.

What does not follow is that we must abandon pre-political values altogether. As Clark aptly notes, a purely epistemic critique of ideology is “toothless” (2024, 264–265); the reason is that a programme which refuses principles while retaining the assumption that value must take the form of principles has nothing left with which to say what would be better. And the commitment to a politics divorced from evaluative aims is itself an evaluative commitment. To insist on seeing political reality as it is, i.e., to refuse illusion and resist self-deception, is already to hold a value, and a demanding one at that. Realism’s animating commitment to staying true to life thus proves pre-political in precisely the sense it set out to deny; it substitutes a different content for the pre-political commitment rather than dispensing with one. Nor is it straightforwardly reality-driven, since confining political theory to stability and institutional arrangement presupposes that these constitute the reality of politics rather than establishing it. The question, I argue, is therefore not whether political theorising ought to begin from values, but which values, held with what disposition, and whether they are openly avowed or concealed.

3.2.3 The wrong target?

The realist charge against Rawls is that he installs a comprehensive morality at the foundation of politics and derives political principles from it. I argue that this amounts to constructing a straw man. What Rawls takes from Kant's moral theory is a conception of persons as reasonable and rational, that is, as having the moral and epistemic powers that make them able and willing to propose fair terms and to recognise the burdens of judgement (see 2.4.1, 2.5.2). He does not, however, carry Kant's deontological ethics over into political conduct. On the contrary, he argues that political liberalism does not adopt Kant's constitutive autonomy and that transcendental idealism plays no part in how the political conception is organised and set out (1996, 99–100).

The moral powers Rawls ascribes to citizens are capacities (1996, 101–105) rather than fixed commitments. Likewise, the Rawlsian sense of justice carries no substantive content and specifies no particular conception of what justice requires. Instead, it supplies motivation – or, in the virtue-ethical terms adopted in this study, a disposition – to seek terms that others could reasonably accept (*ibid.*, 103–104). As Rawls writes, “The worth of citizens in a well-ordered society is always equal because in such a society everyone is assumed to comply with just institutions and to fulfill their duties and obligations moved, when appropriate, by a sufficiently strong sense of justice” (*ibid.*, 280, fn. 16).

Morality is relevant here only insofar as it moves citizens to act in accordance with the principles of fair cooperation. Rawls's sense of justice is not an allegiance to substantive moral truth but a shared recognition of the value of justice as such, and of political cooperation as the means by which a just society is created (1996, 103–104, 127). This, for Rawls, is what makes politics possible as distinct from coercion, and it is all that morality as the sense of justice is asked to do.

The obligation that follows is therefore political rather than comprehensive. The political duty of civility is to explain how the policies one supports might be sustained by the political values of public reason (1996, 217). Rawls does not argue that obligations arising within a comprehensive doctrine are to extend into the public sphere; quite the contrary, he holds that reasonable persons can tell where the line between the two falls, and that the political conception is affirmed alongside a comprehensive doctrine rather than in place of it (*ibid.*, 97–98). In short, what Rawls carries over from Kant is rationality of reasoning, not moral duty.

Realists therefore miss the target in criticising Rawls's pre-political morality, since Rawls, if the reading in 2.2 is right, is a conceptual moralist. For him, political life should respect the reasonable value commitments of citizens while recognising that no one holds a monopoly on moral truth. Citizens may recognise

one another's doctrines as reasonable without being able to recognise them all as true, because there is no shared public basis for deciding which of them is (1996, 128–129). He is not, then, the substantive moralist that realists take him to be.

This is not to say that every aspect of the realist critique is misplaced. Rawls does give justice priority over legitimacy, presupposes that citizens possess the requisite moral and epistemic powers, and does not welcome unreasonable doctrines to the table. Moreover, Geuss's central objection – that abstracting from power is itself an ideological move – remains largely unanswered (2.6.2). Even so, the realist critique is calibrated against substantive moralism, which derives political principles from the content of a prior moral doctrine and then imposes them on politics. The critique thus loses its force against Rawls, because one cannot expel moral truths from the pre-political domain when they were never installed there in the first place.

3.3 THE DIFFICULTIES OF EXCLUDING PRE-POLITICAL MORALITY

3.3.1 Metaethical commitments of realism

Above, I outlined the realist contention that moral norms are the product of historical and social processes internalised by individuals. However, this claim alone is insufficient to justify the rejection of pre-political morality, as the same can be said of numerous other beliefs that likewise emerge from and adapt to specific historical developments and social conventions – such as the view that democracy is the only legitimate form of government, that private property constitutes a political right, or that decentralised governance is preferable to centralised authority. We therefore need to look more closely at how realists understand ethics by examining their metaethical commitments, which help explain and support their central claim that moral norms have no objective basis.

The metaethical commitments of political realists remain largely implicit, as they are not addressed explicitly by realists themselves. This appears to be primarily because realists regard ethics and morality as forms of social facts, akin to any other. For realists, therefore, there is no necessity to treat ethical or moral values as occupying a distinct or autonomous domain. The distinction between facts and values has been identified as the source of a fundamental error made by moralists (Geuss 2008, 16). Geuss extends this critique of the fact-value dichotomy to encompass further oppositions – namely, between Is and Ought, Theory and Philosophy, Descriptive and Normative – arguing that such distinctions are of limited utility in political thought.

“They condemn themselves to certain forms of ignorance and illusion, and introduce into their cognitive and political practice a rigidity and deformation it need not have. In some contexts a relative distinction between “the facts” and human valuations of those facts (or ‘norms’) might be perfectly useful, but the division makes sense only relative to the context, and can’t be extracted from that context, promoted, and declared to have absolute standing... So rather than talking at great length and to no clear purpose about the Is/Ought distinction in general, I would like to proceed indirectly by inviting the reader to see how much more interesting the political world seems to be, and how much more one can come to learn and understand about it, if one relaxes the straightjacket and simply ignores this purported distinction.” (2008, 16–17)

Adrian Kreutz notes, however, that such metanormative foundations of realism are insufficiently distinct to substantiate its status as a discrete mode of thought. Rather, it constitutes a form of “modernist project in this spirit of disenchantment” (2023, 23), characterised by metanormative agnosticism (*ibid.*, 6), including in the realm of metaethics. Nevertheless, an agnostic stance towards metanormativity bears comparable significance for understanding the realist project, as it too reflects a specific choice of foundational premise that, in turn, shapes one’s conception of the task of political philosophy. Let us, therefore, examine in detail the implications of realism’s rejection of the fact-value dichotomy.

By rejecting the fact-value dichotomy, realism suggests that critique can recognise wrongdoing purely through observation. Consider normative behaviourism (Floyd 2011, 2017, 2020, 2023a, 2023b), a sub-branch of realism, which evaluates political principles based on how people actually respond to them,⁹ and its two claims: “[people’s] actions tell us something of their character, and in turn what they really value” (2023b, 495); “Principles are justified for people only in virtue of facts about them” (*ibid.*, 493). In other words, normative behaviourism advocates the formulation of normative standards on the basis of observable responses – especially by examining instances of crime and insurrection as indicators of disaffection or illegitimacy. Here, normativity is not externally imposed but emerges from within the social and practical lives of agents. What is considered the right course of action is thus inferred from those principles and practices that have previously yielded non-harmful or stable outcomes. Floyd argues that descriptive facts about the world and normative

⁹ Normative behaviourism and radical realism share a common ground in their focus on empirical evidence, or fact-centrism (Rossi 2023).

values – what ought to be – are not distinct; rather, our values are embedded within the social realities we observe. Consequently, in this view, we should not regard values as external standards imposed from above, and although we might imagine a better world, such ideal visions offer little practical guidance for action within the constraints of current reality.

The rejection of the fact-value dichotomy makes such an approach possible, but it also means its prescriptive power cannot go beyond what can be justified through practical outcomes, as these are evaluated based on past experiences. As such, for a realist prescription, the constraints of reality emerge from our knowledge of the past, which we then use to predict the possible outcomes of the prescription. After all, reality is what we know today about our past. Thus, if we ground morality merely in our interpretation of past facts about our values, we risk limiting our ethical framework to a retrospective analysis, potentially neglecting forward-looking moral considerations that could address ever-evolving social challenges. Furthermore, this conflation of facts and values reduces values to a purely instrumental, goal-oriented motivational category. As Erman and Möller (2023a) also observe in their critique of normative behaviourism, the approach is fundamentally rooted in utilitarian logic, wherein the evaluative standard is a quantifiable outcome – specifically, a society characterised by minimal dissatisfaction.

To better understand how realists conflate values and facts, it is useful to distinguish between value and worth, following Andrew Collier (1999). Collier, a critical realist, like political realists, characterises values as formed through social interactions, and argues that, by observing such interactions, we can identify formal criteria for discerning values. Value is a product of evaluation and thus contingent upon the valuing agent; in this sense, values can be treated as facts. Moreover, when viewed through the distinction between intrinsic and instrumental values, facts are seen as possessing instrumental value – that is, they are facts for something, or means to ends.

However, critical realism cautions against committing the epistemic fallacy, which consists in reducing ontology to empirical evidence (Bhaskar 2008, 26). Therefore, equating values with facts risks collapsing what ought to be a distinction. Worth, by contrast, denotes something intrinsically valuable, which does not depend on recognition by a valuing agent (Collier 1999). Because it is intrinsic, it comes before both moral and non-moral values, since we must first see what is worthwhile before we can make meaningful evaluative claims (Bhaskar 2008, 59). In this sense, worth can exist independently of values, but values require worth to be intelligible.

Values are the outcome of evaluation and play a functional role in decision-making, whereas worth offers a more flexible reference point for such evaluation. Flexibility here refers to relational activation, whereby worth becomes salient in specific contexts. Collier, for example, discusses a dog that is dangerous to strangers but not to its owner or to a telegraph pole (1999, 56). In short, values require worth to be meaningful. Without worth, values become merely instrumental or functionalist, serving as tools or plans for achieving ends, and allowing only for scrutiny of internal consistency rather than for the fundamental normative critique that political realists claim to engage in (for arguments against functionalism, see [Erman and Möller 2023b](#)).

Consider, for instance, advocacy for peace during wartime. Even where there is broad support for ending a war, this support can fracture once its real costs become explicit, such as ceding territory to the aggressor. This is illustrated by European public opinion on Russia's war against Ukraine in early 2024, since although most expected the war to end in a negotiated settlement and few believed Ukraine could win, Europeans were divided on what to do, with around two in five wanting Europe to push Ukraine to negotiate and roughly a third preferring to help Ukraine regain its territory (Krastev and Leonard 2024). This suggests that people may prioritise peace over the moral principle that 'might should not make right'. The former is rooted in instrumental value and short-term interests; the latter is an inherently normative claim that aggression is morally indefensible under any circumstances. While one might argue against the cessation of hostilities on instrumental grounds – such as the claim that appeasement legitimises aggression and encourages further violations – equally compelling instrumental arguments can be advanced in favour of peace.

The philosophical question, then, is whether these alternatives should be seen as equally valid. In the end, even when we clearly explain the reasons for each option, our decisions still depend on either instrumental values or intrinsic worth. And choosing one over the other is, at its core, a moral choice.

To reiterate, when realists conflate fact and value, they in effect deny that things can have intrinsic worth. They do so for two interrelated reasons: (1) they evaluate worth (e.g., justice) using the same formal criteria as they do for facts; and (2) they argue that, since agreement on intrinsic worth is unattainable, only facts (i.e., phenomena with empirical bearing, such as observable behaviour) possess normative relevance. However, the core issue is that realists appear to be attacking a straw man (see also 3.2.2, which develops the same argument from a different angle). They reject the notion that politics aims to achieve consensus on fundamental principles (Floyd 2017, 92), despite the fact that contemporary post-Rawlsian theorists have largely moved away from consensus models. Instead,

they increasingly adopt frameworks designed to manage disagreement within pluralistic, agonistic contexts (see chapter 4).

The core of the realist critique of ethics-first views lies in the claim that politics is not about agreements, because agreeing on fundamental issues (whether values or interests) is impossible. I will use Bonnie Honig's argument, which William Galston (2010, 386–387) uses to illustrate this point:

“Those writing from diverse positions—republican, liberal, and communitarian—converge in their assumption that success lies in the elimination from a regime of dissonance, resistance, conflict, or struggle. They confine politics (conceptually and territorially) to the juridical, administrative, or regulative tasks of stabilizing moral and political subjects, building consensus, maintaining agreements, or consolidating communities and identities. They assume that the task of political theory is to resolve institutional questions, to get politics right, over, and done with, to free modern subjects and their sets of arrangements of political conflict and instability.” (Honig [1993] 2023, 2)

As a brief commentary, I contend that realists often conflate *harmonious* living with *harmonised* living. The error here lies in the assumption that any political regime – be it republican, liberal, communitarian, or otherwise – entails a broadly uniform form of life, in which differences are minimised or eliminated to produce a seamless, coherent way of living. For instance, it is presumed that a liberal state necessarily consists of liberal citizens. The challenge of envisaging a stable political order (with a certain overarching worldview) while accommodating foundational pluralism has been addressed by various moralist approaches. Realism, by contrast, has thus far focused primarily on the methodological or technical critique of ethics-first political theory – approaches that begin with idealised constructions of justice or legitimacy – without offering a substantive alternative.

Those realists who do attempt to articulate a positive political vision tend to fall back on the liberal democratic order. Williams (2005, 9), for example, argued that liberalism is the (inevitable) outcome of the conjunction of legitimacy and modernity (see also Hall 2020; Sleat 2013 for discussions of realist liberalism). However, liberal realists remain vague regarding the specific institutional and normative structures of liberal political arrangements, particularly when it comes to the hard problem of managing deep value conflict – such as tensions between a liberal state and citizens whose commitments may incline towards authoritarianism. At the same time, they aim to avoid the perceived flaws of moralist theories, namely utopian abstraction and ideological overreach.

Thus, while realists maintain that political life is characterised by enduring internal conflict, they offer limited concrete guidance for managing such conflict. On the one hand, this is consistent with the realist commitment to describing political life as it actually is, rendering realism more *reactive* than prescriptive. On the other hand, realism finds itself caught between a scepticism towards the possibility of consensus and a reluctance to construct a positive vision – whether in the form of ideal theory or more modest normative frameworks. Realists tend to argue that any foundational set of principles is ultimately infeasible and potentially distorting when used to construct possible political futures.

3.3.2 The realist conception of the political

In the broadest sense, realist scholars emphasise that normative theorising about politics should be realistic regarding its nature. This means limiting the normative work strictly to the process of politics and what emerges from the political sphere (Sleat 2010, 487), and treating anything outside it as noise to be eliminated. William Galston borrows from Elkin to outline the purpose of political action: “the various ways in which we arrive at collective, authoritative decisions in a world in which people legitimately hold different views about the purposes of government and the manner in which it should be carried on” (Elkin 2006, 257 n. 1, as quoted in Galston 2010, 387).

While realists generally agree on emphasising the autonomy of the political sphere (Galston 2010, 390; Rossi 2019, 639), there are quite large disagreements within the family on precisely what constitutes the political sphere (Sleat 2018, 2). For Bernard Williams (2005), political theorists must focus on the “first political question” of “securing of order, protection, safety, trust and the conditions of cooperation” (Williams 2005, 3) through meeting the Basic Legitimation Demand (BLD), “in which the state has to offer a justification of its power to *each subject*” (ibid., 4–9, original emphasis). In turn, Sleat argues that realism should encompass not only the narrow issue of legitimate stability, as highlighted by Williams, but also “political possibility, political agency, compromise, political judgement, political institutions, and political responsibility” (2018, 4). For Raymond Geuss (2008), the central issue in politics is related to power, more specifically, “agency, power, and interests, and the relations among these” (ibid., 25), and theorists must therefore ask “Who <does> what to whom for whose benefit?” (ibid.). Geuss urges us to scrutinise the illusions (ibid., 11) that obscure our understanding of “the way the social, economic, political, etc., institutions actually operate in some society at some given time, and what really does move human beings to act in given circumstances” (ibid., 9).

Political realists are not merely committed to critiquing political philosophy for its detachment from practical politics (Leader Maynard 2022); they are also committed to constructing principles for a “[typically interpretative] understanding of the practice of politics itself” (Rossi and Sleat 2014, 690, emphasis added). Williams shares this emphasis on practice, as the title of his *In the Beginning Was the Deed: Realism and Moralism in Political Argument* (2005) makes clear, since what matters for Williams, as for realists more broadly, is political action and not ideas alone¹⁰.

Within this realist literature, we can broadly distinguish two ways of engaging with political practice, namely the Williamsian approach of ‘sensemaking’ and ideology critique rooted in the Frankfurt School of critical theory. I address each in turn, beginning with Williams.

Williams urges us to “make sense” of politics through self-reflection on how it actually functions, and on how our conceptions of it are shaped by socio-historical context:

“For political projects are essentially conditioned, not just in their background intellectual conditions but as a matter of empirical realism, by their historical circumstances [...] there is no way in which theory can get all the way ahead of practice and reach the final determination of what can make sense in political thought [...] there is good reason to think that its [liberalism’s] continued success, now, may require a better, more Wittgensteinian and, more important, Goethean self-understanding.” (2005, 25–26)

The purpose of “making sense” of politics in this way is to understand how politics can address what Williams calls the “first political question” – namely, achieving security and stability without relying solely on coercion, by ensuring broad acceptance of political rules through the Basic Legitimation Demand (2005, 4–5). On this view, what distinguishes political practice from other forms of social practice is precisely the establishment of standards and norms that ensure the response to the first political question is genuinely political, rather than rooted in the classical realist doctrine of *realpolitik*, where ‘might is right’.

This emphasis on practice over theory is captured in the book’s title, which gives priority to the ‘deed’ over the ‘word’ or the ‘thought’ (see also Rowe 1991). For Williams, we cannot approach deeds by starting from words about them, since

¹⁰ The phrase “In the beginning was the deed” (*Im Anfang war die Tat*) originates in Goethe’s *Faust*. As Williams explains, this ‘primacy of practice’ is not a claim about the primacy of *descriptions* of practice (2005, 24, emphasis added).

our perceptions – and therefore also our descriptions – of reality are invariably distorted:

“There could not be a description of our practices which adequately represented the way in which belief was grounded in practice, if only because the understanding of that relation itself would have to be grounded in practice” [...] “For political projects are essentially conditioned, not just in their background intellectual conditions but as a matter of empirical realism, by their historical circumstances.” (2005, 24–25)

It is for this reason that Williams argues that morally driven political theories, such as communitarianism or liberalism, which attempt to establish certain local value perceptions as universal, lack a justifiable basis for doing so (2005, 24–25). This follows from what Sorgiovanni (2020) calls Wittgensteinian constitutive uncertainty, a metaphysical and, by extension, epistemic indeterminacy about the relationship between ‘what is’ and ‘what is told’. Given this indeterminacy, and given that the social world does not exist independently of human interaction (whether direct or indirect), we might say that we create reality through interaction, even as we perceive the fruits of that creation very differently from one another.

In other words, Williams’s reference to Goethe via Wittgenstein implies that what is merely imagined cannot serve as the foundation for actual political practice, precisely because such ideals are inherently arbitrary. It follows that, before any normative prescription can be made, the description, explanation, and understanding of existing political affairs must first account for their socio-historical contingency.

To illustrate how we might “make sense” of existing political realities from a Williamsian perspective, consider the position of women in society through the lens of the Basic Legitimation Demand (2005, 4–9). This demand requires the state – as the holder of power – not only to provide stability and security, but also to justify the means of doing so beyond mere coercion, aiming for the active and peaceful acceptance of those means by those subjected to them – in this case, women. The question that arises, then, is whether the Basic Legitimation Demand can be satisfied in a context where women are in some way more restricted than men. For Williams, arguments for justification must be intelligible to the women subjected to them (2005, 26–28); where the relevant criteria for justification are met within that context, legitimacy can be considered achieved. This is not to say that our normative work must proceed from such a non-presentist approach to legitimacy – as Williams himself notes, “no political theory, liberal or other, can determine by itself its own application” (ibid., 28).

Rather, it allows us to evaluate the legitimacy of women's position without grounding our point of reference in an ideal version of gender equality that is itself contestable, both temporally and conceptually¹¹.

However, as one might suspect, there is a flaw in this reasoning, since it appears to permit gender inequality as long as everyone accepts it. By merely acknowledging historical contingency without addressing it critically, Williams risks normalising that contingency, thereby favouring the maintenance of the status quo (see also [Philp 2010](#); [Finlayson 2017](#); [Kreutz 2023](#); [Westphal 2023](#); [Cross 2024](#)). Put differently, while Williams's 'sensemaking' approach helps us understand the criteria for legitimacy as they exist in a given time and place, it offers no principles for identifying a misalignment between 'what is' and how the subjects of legitimacy themselves perceive reality. The underlying assumption – that those subjected to coercion possess the capacity to evaluate whether the state's justification is adequate – is therefore flawed: power structures that favour the dominant order can condition women to perceive that order as either inevitable or even in their own interest.

This status-quo bias is precisely what the Geussian radical camp of realists sets out to correct. Rather than taking legitimacy criteria largely at face value, as Williams's sensemaking does, this camp argues that the distortion between the perception and practice of politics must itself be confronted directly – and that the appropriate tool for doing so is ideology critique. Realist ideology critique ([Prinz and Rossi 2017](#); [Rossi and Argenton 2021](#); [Aytac and Rossi 2023](#); [Clark 2024](#)) seeks to identify self-justifying power structures by exposing the ideological sources of the beliefs agents hold – beliefs distorted by the very power mechanisms that shape their formation in the first place ([Geuss 2008](#), 87–88). As Rossi puts it, the realist's role is to unmask the political practices underpinning legitimating narratives that are not what they purport to be ([Rossi 2019](#), 645). In this sense, ideology critique questions the product of sensemaking. Instead of asking only whether the criteria for legitimacy are met, it asks whether those criteria are themselves illusions, created and sustained by those with epistemic power.

3.3.3 The realist search for purely political normativity

As realism emerged in opposition to ethics-first political philosophy, its normative focus is the evaluation of political arrangements and the provision of action-guidance without subordinating politics to moral norms. Realists judge political arrangements by how well they fulfil the primary function of politics,

¹¹ By this I mean that today's ideal may not remain so tomorrow, and our understanding of the relevant concepts will likewise evolve. Social gender, for instance, may lose relevance in the future, potentially dismantling the hierarchies of injustice currently rooted in gender(ism).

namely making binding collective decisions amid persistent disagreement (Burelli 2022). Rather than asking how just, free, or egalitarian an arrangement is, realists ask how binding agreement is achieved at all. The defining feature of realist normativity, in short, is the attempt to avoid norms with pre-political, substantive moral content.

Both radical and moderate realists reject moral norms as the foundation of political action. They part ways, however, on how far this rejection goes. Radical realists deny morality any justificatory role at all, whereas moderate realists allow it a place within politics, as long as it has first been filtered through political reality and reshaped by the features that make politics what it is (Erman and Möller 2025, 27).

If not moral norms, then what does ground realist normativity? Four candidates have been proposed – the moderate realist *conceptual* (Sleat 2014) and *filter* views (Jubb 2019; Sleat 2022), and radical realist *instrumental* and *epistemic* views (Burelli and Destri 2022; Burelli 2022; Aytac and Rossi 2023). Each, on inspection, fails to supply a self-standing alternative to morality.

The *conceptual* view holds that the very idea of politics itself provides the foundation for a distinctive form of political normativity by distinguishing it from mere domination. This understanding of politics as *non-domination* draws on Williams's (2005) argument that grasping the nature of politics requires recognising its inherently evaluative dimensions – akin to so-called 'thick' evaluative concepts such as 'brave' or 'cruel' (see also Nye 2015, 1091), which are not merely descriptive but carry an evaluative component. Politics is seen as valuable for its role in organising the social world, concerned with questions of power, order, and stability, without necessarily addressing the purposes to which these are directed. But as Erman and Möller (2015) argue, distinguishing politics from domination is not enough to ground a comprehensive normative framework, because it would specify what politics is without discussing what kind of politics is desirable or legitimate. Simply put, they argue that conceptual distinction cannot by itself generate substantive norms.

The *filter* view differs from fully non-moral approaches because it accepts that morality has a role in political theory, while denying that political judgements can be derived directly from moral values. Instead, moral principles must be filtered through the realities of politics (Sleat 2022, 474). Political conflict, power relations, and institutional constraints change how moral considerations apply in practice (Jubb 2019, 362). This allows political and moral judgement to come apart, since citizens may have a political duty to accept a legitimate decision even when they believe it is morally wrong (Erman and Möller 2022, 6).

Critics argue, however, that this filter is asked to do more work than it can justify. As Erman and Möller (2025, 37) note, the idea of a political filter assumes a fixed political reality against which moral principles can be adjusted. On this view, personal morality is ‘thinned out’ to fit the values of existing political institutions. However, there is no method to decide to what extent moral principles should be adapted or what counts as politically feasible in terms of adopting new values. The filter view also asks a lot of political actors by expecting them to play different roles in different areas and to justify each role by a different moral standard. Yet it never explains how these standards can fit together in one ‘filtered’ ethical outlook.

The instrumental view tries to avoid morality altogether by defining political norms purely in terms of means-ends reasoning, in which stability is the end, since “the securing of order, protection, safety, trust, and the conditions of cooperation” is the precondition for everything else politics might achieve (Williams 2005, 3). Erman and Möller (2022, 4) critique this by contending that this is not self-sufficient, because the normative weight given to stability depends on a particular conception of what politics is, while that conception is itself contested. In other words, an instrumental framework smuggles in the kind of contestable starting point it wants to avoid.

The epistemic view focuses on factual accuracy and the critique of ideological distortion – using empirically grounded genealogical analysis to expose how power and false consciousness shape political arrangements (Aytac 2022; Burelli and Destri 2022; Raekstad 2021; see also 3.2.2). But as Erman and Möller (2022) argue, epistemic clarity is not sufficient for normative conclusions. Even when the facts are clear – for example, when looking back at the 2003 Iraq invasion – deciding what should have been done, and why, still requires practical moral reasoning that simply having the facts right cannot provide.

None of the four realist proposals, then, succeeds in replacing moral norms with a purely political source of normativity. The reasons for their failure point to two deeper structural costs of attempting to avoid moral premises altogether.

First, both epistemic and instrumental normativity are functional in nature, since they presuppose an end, and ends require a foundation because they cannot create value out of nothing (Leader Maynard and Worsnip 2018, 779). What they require, then, is a substantive conception of what is necessary to sustain a stable yet diverse society in which individuals willingly accept certain binding norms. Yet, how the substantive aims of politics are constructed (or emerge) is not immediately clear. How is the balance determined between stability, material interests, and human values if the primary normative criterion is legitimacy (Sleat 2014; Cross 2020) or, in more modest terms, the capacity to prevent worse

scenarios by employing well-informed knowledge (Bagg 2022), as realists argue? So if stability or legitimacy is to function as a meaningful evaluative standard – rather than an arbitrary stopping point – it must rest on something of intrinsic value. Politics built on instrumental aims alone, without any intrinsic value to anchor them, cannot get off the ground.

Second, factoring out pre-political ethics licenses political action justified by success alone. Radical realism, on this point, imposes no requirement to honour reciprocity, fairness, or human dignity. Burelli and Destri's Machiavellian formulation says that "the facts of political competition tell us that the ability to lie is required to achieve any end" (2022, 409). Divorcing ethics from politics this way reduces individuals to 'political animals', driven solely by rational self-interest and stripped of any profound sense of soul or moral depth, ultimately eroding the foundations of a humane society.

Taken together, unless we are willing to imagine politics entirely devoid of values, some pre-political ethical commitment is necessary – though not sufficient – to direct political action. Without it, supposedly purely political values like stability and legitimacy stop serving any external purpose and become ends in themselves, with politics organised around its own self-perpetuation, prioritising order and control for their own sake, rather than around anything, like freedom, actually worth securing.

This is why the previous section examined the realist ideology critique. While the conceptual, filter, and instrumental approaches remain vague and underdeveloped (Erman and Möller 2025, 27), and lack substantive scholarly examples illustrating their application in political theory, the epistemic approach via ideology critique is realism's most fully worked-out attempt to ground normativity without morality (Finlayson 2017; Rossi 2019; Cross 2022). Examining it in detail showed that, absent moral grounding, ideology critique – and, more broadly, immanent critique as its method – will ultimately fail to fulfil its objective of fostering meaningful social critique and transformation.

3.3.4 Realism and liberal legitimacy

If realists reject morality and justice as the primary evaluative concepts for politics, what do they offer instead? Both radical and moderate realists appeal to legitimacy, but only moderate realists have developed this notion in any detail. For Williams, legitimacy is the central concern for political theorists evaluating political arrangements (2005, 48–49). In modern pluralist societies, he argues, liberal arrangements may be best suited to secure legitimacy by meeting the Basic Legitimation Demand (2005, 9) – though he stresses this is a contingent historical judgement, not a universal claim. This puts Williams at odds with Rawls's

political liberalism, where political power must instead satisfy certain “moral conditions of co-existence under power – conditions in which power can be justly exercised” (ibid.).

For Williams, liberalism emerges as the only viable political order at the intersection of modernity and legitimacy, where modernity denotes the erosion of shared moral or religious frameworks, while legitimacy concerns the requirement to justify the exercise of power. On this view, liberalism does not require agreement on foundational moral truths; it merely provides a legitimacy-grounded framework within which disagreement can be *managed*. Williams thus agrees with Rawls that liberalism suits societies marked by deep value pluralism, but the two diverge on what sustains that order, since where Rawls treats moral conditions as foundational (2005, 1, 19), Williams substitutes legitimacy. His formula – “LEG [Legitimacy] + Modernity = Liberalism” (2005, 9) – casts legitimacy as what constitutes the polity, echoing Geuss’s claim that “understanding legitimacy is required for understanding how society behaves politically” (2008, 34, 36). In short, realists prioritise legitimacy, understood as a political system’s capacity to provide stability, order, and justification in terms meaningful to those subject to its authority (Sleat 2014), over the prior question of what makes a society good.

However, this substitution fails for at least three reasons.

First, legitimacy cannot escape the historical contingency it inherits. Realists have used immanent critique to interrogate the self-legitimising tendencies of political orders, including liberal ones (Aytac 2022; Aytac and Rossi 2023), but such critique remains bound by whatever standards of legitimacy currently prevail. Without an independent standard of comparison, present arrangements risk being judged adequate simply because they compare favourably to the past – mistaking ‘better than yesterday’ for ‘good enough today’. As Sleat puts it, “legitimacy does not stand above the fray of political contestation but is itself deeply immersed in it ... the process of legitimation ... is part of the very struggle for power and influence that is politics” (2014, 330). In other words, realism’s commitment to political reality has a cost – it cannot provide standards that go beyond present knowledge because it tests only whether something fits with what we already know about politics. Realist positive theory would thus amount to merely a parallel construction of the present.

Second, Williams’s legitimacy cannot ground rights without implicitly relying on the morality it wants to exclude. Williams identifies torture, assault, and arbitrary power as clear human rights violations (2005, 19), but the realist framework cannot explain why these are violations without smuggling in ethical content. Arbitrary power might be difficult to define, but acts such as torture and

assault are clearly wrong because they involve cruelty. Williams is not able to escape the problem he identifies in the Kantian rights framework, arguing that the chain “rights-principles-moral discourse-person-autonomy-rights” (ibid., 21) is circular. However, Williams’s own view of rights based only on legitimacy is also problematic, because if legitimacy is the only standard for evaluating political power, then it may be able to defend its legitimacy within both just and unjust arrangements. Thus, without some moral standard we have no way to tell the difference between genuine legitimacy and a system simply constructing the veil of legitimacy to defend its own power.

Consider Williams’s own example of gender discrimination. He argues that claims to gender equality are politically legitimate only when their historical context makes them intelligible, “unless this ‘sense’ itself results from a self-justifying abuse of power” (2005, 26–27). But we cannot take this idea of ‘self-justifying abuse of power’ seriously without some substantive idea of justice to distinguish between an ‘actual’ historical context and a ‘constructed’ distortion created by those in power to make their rule appear legitimate. So, whenever the LEG-framework is applied to questions of rights, historical injustice, or what counts as ‘arbitrary’, it either falls short or brings back the very moral commitments it seeks to avoid.

3.4 CONCLUDING REMARKS: POLITICAL REALISM AND ITS REDUCTIONS

Political realism charges the Kantian-Rawlsian tradition of political philosophy with abstraction and dogmatism, arguing that it puts morality before politics, since pre-political moral norms, on this view, not only abstract from the substance of political life but may also obscure the operation and abuse of power. However, I argue that part of this charge rests on a misconstrual of Rawls’s moralism, taking it to be substantively moral when it is instead conceptual, so that the comprehensive morality realists set out to dislodge is not the one political liberalism installs.

In addition, realism’s own positive alternative faces difficulties that hold independently of whether Rawls is its rightful target. Realists argue instead that political philosophy should take politics as constituting an autonomous and distinct sphere, governed by its own principles, such as stability, legitimacy, and power, and that moral principles should not be used to evaluate political arrangements, since morality is not an expression of objective truth but a historically and socially contingent product. In the absence of an ‘objectively correct’ conception of moral values such as justice or well-being, the realist task

is therefore not to ask how politics might improve citizens' lives, but rather to assess political arrangements in terms of their legitimacy – that is, the extent to which power can justify itself without resorting to self-justificatory narratives.

In seeking to focus on what arises strictly from the political domain, realists have developed several alternative approaches to normativity, intended as correctives to moralism. Moderate realists favour conceptions that treat morality as internal to politics, but only in a contextual and derivative way – as in the conceptual and filter views. Radical realists, by contrast, aim at a fully 'pure' form of political normativity, derived either from practical necessity (the instrumental view) or from the critique of ideological distortion (the epistemic view).

The foundational belief of political realism is that politics can be conducted independently of morality – that political norms, such as those concerning legitimate authority, can be evaluated without recourse to moral norms. Williams (2005) holds that if we abandon moral ideals rooted in modernity, and focus instead on the possibility of a stable yet non-coercive society, we arrive at a liberal order regardless. On this view, we need not posit a moral conception of justice as the aim of politics; we need only relinquish moral premises and let legitimacy do the work, since moral norms are said to emerge from within a legitimate order, rather than to precede it. Radical realists go further still, urging us to scrutinise the failures of such legitimacy directly – by examining the self-justificatory narratives of power holders, which rely on moral norms that are not genuinely legitimate, but merely dogmatic.

Yet realism's avoidance of pre-political justice and other moral ideas faces a basic problem. As Cohen (2003) argues, facts cannot tell us what we ought to do; they can guide us only if some normative principle explains why those facts should count as reasons for action. In other words, action guidance presupposes meaning (and thus, values) behind it, but mere facts have no meaning in themselves. Realism's reliance on 'mere political facts' must therefore already depend on the kind of value-based normativity it claims to avoid.

This chapter has traced three such difficulties, corresponding to realism's three central moves, namely prioritising legitimacy as the primary political objective; developing a 'purely' political form of normativity to sustain it; and employing ideology critique to expose the self-justifications of power holders. However, each runs into the same underlying problem.

First, separating legitimacy from morality leaves arbitrary power undefined. Realists want to distinguish legitimate from arbitrary power, but lacking moral foundations, they struggle to specify what makes power arbitrary in the first place. Williams condemns acts like torture, for instance, but avoids appealing to

universal moral grounds, relying instead on context-specific political justification, leaving no stable basis from which to mount a critique. Realist tools such as idealised models or immanent critique fare no better, since, tethered to existing norms, they risk reinforcing the status quo, or mistaking today's values for normative ideals. And the circularity realists impute to moral theorists, who derive rights from concepts like personhood or autonomy, reappears in their own accounts, which still must condemn torture somehow, without any shared moral commitment to appeal to. Without moral grounding, in short, realism lacks the resources either to justify rights or to meaningfully criticise their abuse – leaving its account of legitimacy both limited and unstable. The very choice to treat power and legitimacy as the relevant political concepts is, moreover, already a normative one – a first sign that description and evaluation cannot be separated as cleanly as realism supposes.

Second, purely political normativity – whether conceptual, filter-based, instrumental, or epistemic – cannot escape value-laden presuppositions either. Each of these approaches grounds political decision-making in empirically informed reasoning. But if legitimate political decisions are based only on empirical facts, while moral values are excluded, politics risks becoming simply descriptive rather than a project aimed at real progress. The problem is the idea of progress itself, because it suggests moving towards something better, but this requires a judgement about what counts as better. Yet realism cannot make this judgement without introducing the moral standards it claims to avoid.

Third, ideology critique, in both of its versions, cannot provide the standard needed to evaluate the problems it identifies. Geuss's version sees morality as ideological when moral norms become fixed and used as tools of power (2008, 42). While this shows how comprehensive moral doctrines can be used to justify hierarchy, it pays less attention to the economic and cultural ideas that can also be used to justify power. Aytac and Rossi's (2023) version focuses more broadly on motivated reasoning rather than moral norms specifically. But identifying a distortion does not tell us what to do about it. Because this approach relies on empirical observation, it cannot tell us which forms of structural injustice matter or how we should respond. This is because such injustices may not be immediately recognisable through observation alone. As Aytac (2022) rightly argues, people may accept stories that justify the dominant power because these narratives have become part of their social identity, and if people have no way of imagining real alternatives, they have little reason to reject these stories, even when they cause real harm.

That said, on my view, realist ideology critique gets at least one thing right, namely that it shows how morality can be used as a discursive means through which power justifies itself.

Taken together, these problems reveal a misunderstanding of the nature of politics itself. By focusing on empirical facts and treating them as objective, realists tend to see political disagreement as a conflict over how best to solve a shared problem. This is because, from their perspective, the problem appears to be the same for everyone. As a result, realism manages the effects of political conflict rather than addressing its deeper sources. Consider debates about the public display of religious symbols, such as the veil (see also 6.6.5). Realism may treat this as a problem of visible pluralism, in that people acknowledge that others hold different beliefs, but they object to the visible presence of religion in public life. However, the deeper issue is not just that people disagree about what belongs in public spaces. It is that the same symbol can mean very different things to different people (Scott 2010), e.g., for some, the veil stands for personal freedom and religious identity, while for others, it is a sign of patriarchal oppression (Abbas 2023). These reasons emerge from different ways of seeing the world, which shape how people conceptualise the problem at hand in the first place, which is something that realism struggles to address.

To conclude, I have made two criticisms of realism while also identifying one of its insights. First, realists misread Rawls by assuming that political liberalism requires acceptance of a substantive moral ideal as the standard for evaluating political arrangements. But Rawls's moralism is conceptual rather than substantive, in that he assumes only that citizens possess a sense of justice sufficient to motivate cooperation, not that politics should aim for any particular conception of justice or moral truth. Second, purely political normativity separates politics from the value commitments that actually shape political life. In practice, citizens do not distinguish the political from their foundational commitments in the way realists envisage. By ignoring this, realists undermine their own claim to offer a true-to-life approach.

But realism nevertheless reminds us that political theory must attend to illusion-making, ideological distortion, and self-justification as problems in their own right, since such distortions do not reveal themselves through simple observation alone and therefore require epistemic scrutiny. That said, realism does so at the cost of allowing stability and legitimacy to function as the standards by which political arrangements are judged, while leaving other equally valuable, if not crucial, features of a healthy political life at the margins. By failing to interrogate its own commitments to stability and legitimacy, and without explicitly

recognising them as values, realism settles in advance which substantive elements of social life political theory should prioritise.

4. POLITICAL AGONISM

4.1 WHAT IS POLITICAL AGONISM?

This chapter turns to political agonism as a second challenge to Rawlsian political moralism, alongside the realist challenge examined in the previous chapter. Agonists conceive of democracy as a condition of persistent and respectful contestation between adversaries who grant one another legitimacy. On this view, the viability of democracy depends directly on agonistic struggle itself, since conflict is not considered a defect to be remedied but rather part of what democratic life consists in and, at times, as Bonnie Honig puts it, one of its pleasures ([1993] 2023, ix).

Bonnie Honig is also the theorist most commonly associated with the emergence of agonism in political theory through her *Political Theory and the Displacement of Politics*, first published in 1993 ([1993] 2023, ix). Subsequent contributions followed, most notably Chantal Mouffe's (1999) theorisation of agonistic pluralism, now widely regarded as the cornerstone of political agonism, as well as William Connolly's ([1991] 2002, 1995, 2005, 2011) ethos of agonistic respect and critical responsiveness. There are also notable works that employ agonism to address more specific aspects of politics, such as James Tully's (2008a, 2008b) conception of agonistic struggles for independence within decolonial and civic forms of constitutional engagement, Patchen Markell's (2003) agonist critique of communitarian conceptions of recognition, and Aletta Norval's (2007) extension of the post-structuralist democratic theory of Mouffe and Laclau, with a particular emphasis on subjectivity, while some other contributions focus more on the individual dimension of self-formation, such as David Owen's perfectionist (existential) agonism (1996, 2002, 2018), but this study will not go into detail on these authors.

Although agonist theorists differ in their emphasis on particular aspects of agonistic democracy, they nevertheless converge on the view that conflict is an inevitable part of social life, while maintaining that the evaluative criteria for political action must ultimately be grounded in legitimacy (Wenman 2013, 6). In doing so, agonism aligns with the political realist critique of liberal theorising, namely that its idealist projects are overly abstract and detached from the contingencies of the world, as they focus narrowly on questions such as 'what is justice?' (ibid.; Honig and Stears 2014, 131).

Mark Wenman offers a comprehensive characterisation of agonism that warrants being quoted in full:

“Unlike most normative political theory in the Western canon, theorists of agonistic democracy do not seek to ground their respective visions of politics in substantive accounts of human nature, in teleological assumptions about the good life or the course of history, or in consequentialist theories of morality. Nor do they share the currently predominant liberal view that agreement on constitutional essentials can be secured through recourse to deontological procedures or a ‘public use of reason’, which purportedly brackets off fundamental disagreements between competing ‘comprehensive doctrines’ or conceptions of the good” (Wenman 2013, 6)

Importantly, agonists contend that liberalism not only abstracts from reality and relies upon utopian assumptions, but also, and more significantly, conceives of the political world as a struggle between fixed properties that merely require correct calibration. By contrast, agonism recognises that properties are always in flux and that new ones continually emerge from what Wenman describes as the “co-originality of the constituent and constituted powers” (2013, 15). We ought to remember, however, that it is not liberalism which stands as the polar opposite of agonism, but rather any form of fundamentalism, whether religious, secular or even democratic (Connolly 2005, 136). Put simply, agonists share a diagnosis of liberalism as a project that seeks to depoliticise foundational differences (and, by extension, substantive pluralism), which, in the agonists’ view, cannot be marginalised, as it is intrinsically valuable to democracy (Fossen 2008, 378). They do not consider it possible for political theory to devise procedures or institutional designs that would neatly define democracy or eliminate structural injustices through attempts to smooth over disagreements and achieve an arrangement-based consensus.

Politics should be understood as being in a constant state of flux because, as Wenman – drawing on Arendt’s insight – points out, “genuine political innovation is always something exogenous to process and disruptive of process” (2013, 102). For Arendt, people constantly bring something new into public life by acting and interacting with others, and because people can always introduce new ideas and possibilities, political life will inevitably be disrupted in unexpected ways; we should not, therefore, attempt to confine such exogenous disruption within rigid political procedures. As she observes in *The Human Condition*:

“it is not the beginning of something but of somebody, who is a beginner himself... It is in the nature of beginning that something new is started which cannot be expected from whatever may have happened before.... The fact that man is capable of action means that the unexpected can be

expected from him, that he is able to perform what is infinitely improbable.” (Arendt [1958] 1998, 177–178)

Such disruption, by introducing what is new and unforeseeable, is essential to politics, for it continually reopens a space of freedom. When people come together in public through non-anonymous action, their activity does not merely generate disorder; it creates the conditions for genuine political innovation, which cannot be fully produced by established political processes, since such processes are designed to reproduce an existing order. What innovation requires is a publicly open togetherness, or action with a name, a ‘who’ attached to it (Arendt [1958] 1998, 180).

Arendt emphasises that when we act, we do not simply perform deeds – we reveal who we are. This disclosure takes place most authentically in the presence of others, without hostility or instrumental calculation, in what she calls “sheer human togetherness” (Arendt [1958] 1998, 180). Such action requires both sincerity and a willingness to take risks. For Arendt, true political life occurs when people appear before one another without concealing themselves. Thus, her notion of disruption is inscribed within the human condition itself. Free politics, on her account, presupposes free individuals – and freedom here is not only granted to others but exercised, by showing oneself. Public space is the shared realm of visibility in which such showing becomes possible. Where it is absent, action loses its meaning and the actor drifts into marginality and isolation (ibid., 180–181).

Agonism shares certain affinities with contemporary political realism. Both traditions view social order as historically situated, inherently diverse, and therefore contingent, rather than the outcome of neutral procedures or universally shared values. They recognise that conflict and disagreement are intrinsic to political life, and they reject the Rawlsian assumption that political consensus can be achieved through rational deliberation under idealised conditions (Williams 2005; Geuss 2008; Selg 2012, 84–85). But agonism goes further, insisting that foundational differences should not be eliminated, whereas realists focus more modestly on ensuring legitimate stability or examining the sources of political and moral illusion. In short, both realists and agonists oppose the moralist (liberal) view that politics can be grounded in universal moral principles independent of power and conflict. The connection between realism and agonism is neatly outlined by William Galston in one of the earlier broader discussions of political realism:

“Honig acknowledges that ‘[t]he perpetuity of conflict is not easy to celebrate’, and she does not do so. It is rather to say that politics is always and everywhere a tension between the drive for and goods of

stabilization and consensus, on the one hand, and the drive for and goods of destabilization and conflict. To recognize this enduring duality is to acknowledge that the maxims of law, economics, and morality will never displace the need for political decisions, informed but undetermined by any intellectual discipline.” (Galston 2010, 396)

Yet agonism and realism differ in their respective normative orientations, as agonism seeks a more affirmative and participatory engagement with politics, whereas realists may adopt either a primarily critical and diagnostic stance (as in the radical, Geussian strand) or a *modus vivendi* approach that emphasises order and stability (as in the more moderate, Williamsian version) (Honig and Stears 2014). In agonism, political identities are understood as discursively constructed, fluid, and formed through relations of difference and exclusion, with particular emphasis on the affective and symbolic dimensions of political life (Mouffe 2013), whereas realism operates within a more empirical and historically informed framework, presuming relatively stable political interests, identities, and constraints, and focusing on the ideologically compromised and context-bound nature of real-world political action (Rossi and Sleat 2014). Most agonist theorists accordingly advance an emancipatory project¹² aimed at fostering pluralism and transforming antagonistic conflicts into agonistic engagements (Mouffe 2013), whereas realists prioritise feasibility and the constraints of political practice (Cozzaglio and Favara 2022).

With this brief introduction in place, we can now move on to discuss the three forms of agonism developed by Honig, Connolly, and Mouffe (4.2–4.4), followed by an overview of how these agonist thinkers conceive of morality (4.5). These theorists have been selected because each develops a distinct foundational dimension of agonism. Others (e.g., Tully, Markell, Norval, Owen) engage the tradition more selectively, taking up particular aspects of it rather than supplying its foundations.

4.2 BONNIE HONIG’S JOY OF AGONISM

Bonnie Honig has been described as a leading initiator of agonistic thinking in political theory during the mid-1990s (Honig [1993] 2023, ix; Honig and Cole 2021). Honig contrasts her approach with political theories that seek to stabilise society, most notably the political liberalism of Rawls (Kantian) and his communitarian critic Sandel, where theory aims to establish principles for improving political

¹² With the notable exception of David Owen, who has been characterised as a perfectionist agonist (Fossen 2008).

institutions in order to bring political conflict to an end ([1993] 2023, 2). As Honig observes,

“Most political theorists are hostile to the disruptions of politics. Those writing from diverse positions—republican, liberal, and communitarian—converge in their assumption that success lies in the elimination from a regime of dissonance, resistance, conflict, or struggle. They confine politics (conceptually and territorially) to the juridical, administrative, or regulative tasks of stabilizing moral and political subjects, building consensus, maintaining agreements, or consolidating communities and identities. They assume that the task of political theory is to resolve institutional questions, to get politics right, over, and done with, to free modern subjects and their sets of arrangements of political conflict and instability.” ([1993] 2023, 2)

In contrast, Honig argues that the closure of political conflict is not only impossible but also undesirable, as it inevitably produces what she calls “remainders” ([1993] 2023, 213 n1), a term she borrows from Bernard Williams. Remainders are “misfits or ne’er-do-wells who would likely suffer under the proposed arrangements, but whose suffering would not rise to the threshold of audibility or visibility” (Honig and Cole 2021, 351); instead, they are depoliticised in the name of stability and order (Honig [1993] 2023, 5). In contrast, Honig’s agonism seeks to provide space for those ‘remainders’ who are excluded by the dominant discourse to reclaim their voices. In this sense, it constitutes an emancipatory project.

4.2.1 Virtù, not morals

Honig’s project develops two conceptual frameworks to distinguish between stability-seeking approaches to politics and disruptive approaches that resist the aspiration to permanent closure. Somewhat confusingly, she labels the former – the approach she critiques – as the “virtue theories of politics”, while describing her own as the “virtù theory of politics” ([1993] 2023, 2). A clarification is therefore in order.

For Honig, the term virtue does not refer to virtues of character. Rather, Honig’s virtue denotes what the present study understands as morality – a system of thought that seeks to stabilise political orders through establishing particular truths ([1993] 2023, 2), while this study distinguishes (virtue) ethics as an epistemically oriented mode of ethical reasoning (1.4; chapters 5 and 6). In this respect, Honig employs virtue in a manner broadly analogous to the way Bernard Williams, Nietzsche, and this study use the concept of morality: on her account,

both contractual and procedural liberals, as well as communitarians, pursue such closure, albeit through different means ([1993] 2023, 2).

Thus, to avoid confusion, whenever Honig uses the term virtue in this sense, I will add [moral], while the term denoting the opposite of morality, *virtù*, will remain unchanged.

The term *virtù* originates with Machiavelli, for whom *virtù* signifies political agency – or, more precisely, political excellence in the capacity to act decisively, even disruptively, under conditions of uncertainty, with the aim of “founding and maintaining a lasting regime” ([1993] 2023, 68). However, while Honig accepts Machiavelli’s notion of political glory, she departs from his understanding of its purpose. In developing her own conception of *virtù*, she instead engages primarily with Nietzsche and Arendt. In brief, Honig’s *virtù* concerns resistance to the dictates of the dominant canon; its purpose is to serve as the disruptive force that destabilises prevailing arrangements. Honig cites Nietzsche, who observes that “the man who obeys the dictates of the moralist” ([1993] 2023, 68) “can achieve greatness in this way” but is also “cut off from the most beautiful fortuities of his soul, and from all further instruction” (Nietzsche [1882] 1974 §305, via Honig [1993] 2023, 68, original emphasis). As such, *virtù* is a form of excellence; yet, unlike the Machiavellian, power-oriented conception that seeks closure, its excellence lies in the “strategic disruption of the impositional orderings of the herd and the alternative construction of the self as a work of art” ([1993] 2023, 69). Thus, Honig transforms Machiavelli’s *virtù* aimed at regime stability into a form of creative disruption that resists conformity and challenges established political and moral norms.

With this clarification in place, let us turn to Honig’s own account of the ‘virtue [moral] theories of politics’:

“Virtue theorists of politics assume that the world and the self are not resistant to, but only enabled and completed by, their favored conceptions of order and subjectivity. This assumption undergirds their belief that modern disenchantment, alienation, pain, and cruelty would be diminished if only we adopted their principles of right, established just institutions whose fairness is ascertainable from a particular (rational) perspective, or yielded to the truth of membership in a wider community of meaning and value. In short, each of the virtue theorists examined here believes, mistakenly, that his own theory soothes or resolves the dissonances other theories cause. Each yearns for closure and each looks to politics, rightly understood, to provide and maintain it.” ([1993] 2023, 3).

Let us unpack this. Honig argues that the virtue [moral] theories (represented in her discussions of Kant, Rawls, and Sandel below) share a form of optimism that none can ultimately fulfil. Their mistake is to assume that both the world and the self naturally fit the political order they advocate. On this view, the self is never fundamentally at odds with the political order proposed for it. Honig thus argues that the virtue [moral] theories rest on the simplistic assumption that, if we adopt the right principles, the tension between the individual and society can, in principle, be overcome. This assumption explains why virtue [moral] theories regard many modern problems (e.g., disenchantment) as largely resolvable. For instance, a Rawlsian expects a just political order to reduce persistent conflict over the distribution of resources, a communitarian hopes that a shared sense of meaning will overcome alienation, while a deliberative democrat considers public reasoning the remedy for alleviating disagreement. Honig draws our attention to the fact that each theory believes (mistakenly, in her view) that it resolves the tensions created by its rivals by identifying the blind spots of competing theories while remaining confident that its own framework can achieve a stable political order. Honig argues that this confidence comes from seeking a clear solution, grounded in the belief that politics, done properly, can build a stable order where society's deepest conflicts are more or less settled, so that they do not have to be constantly negotiated.

In contrast, Honig considers political adversaries not as obstacles to be overcome on the path to eventual consensus but as permanent and equal participants in an ongoing process of contestation. Moreover, in my reading of Honig, they exemplify what I term 'adversarial sparring' (see section 4.2.3), the idea that contestation is not merely unavoidable but also constructive. Adversaries challenge each other's arguments, question assumptions, and point out blind spots. This helps both sides get clearer about their views and why they matter to them. Disagreement then becomes a way of learning from each other, not a struggle to be won.

4.2.2 How morals displace politics

Honig developed her account of agonism in *Political Theory and the Displacement of Politics*, first published in 1993. In this work, "displacement" refers to the theoretical suppression or domestication of political contestation in favour of order ([1993] 2023, 2). To explicate this displacement, Honig turns to three political philosophers – Kant, Rawls, and Sandel – examining the inconsistencies between their stated assumptions and their intended aims. In particular, Honig contends that both liberal and communitarian theorising attempt to resolve political conflict, ultimately fail to do so, and, in the process, obscure it. Kant appears here as the chief proponent of human rationality and reasoning, and this

conception became one of the central themes in the liberalism–communitarianism (e.g., Rawls–Sandel) debate that shaped political philosophy during the 1980s. This debate, in turn, became the focus of Honig’s subsequent critique and gave rise to the agonistic alternative.

Honig’s view of Kant

Honig’s reading of Kant is somewhat unconventional: she argues that Kant’s concept of respect for persons is not actually respect for persons tout court, but rather respect for the moral law, which is then transferred to persons as their potential for conversion to moral worthiness ([1993] 2023, 18) with the ultimate ambition of reaching moral progress in humans (ibid., 19). Honig also notes that it is not respect for persons but respect for progress that justifies Kant’s political orderings (ibid.).

The reliance on rationality and a belief in progress is significant because it yields a conception of humans as benevolent beings – willing and able to improve their moral condition. While Honig critiques Kant for his view that moral ends are objective and preconfigured, merely awaiting “ruptures of reason and its ensuing rapprochements” to be realised ([1993] 2023, 20), she nevertheless appears to share Kant’s optimism regarding human nature. After all, agonism, at the very least, presupposes continuous self-correction (ibid., 204) in order to avoid collapsing under its own nonconformity and emphasis on persistent instability. This, however, also presupposes rationality, albeit without objectivity. Thus, she critiques Kant for envisioning a determinate outcome and stable conditions, rather than the process of rationality and reasoning itself.

Honig’s view of Rawls

Honig’s ([1993] 2023, 126–161) reading of Rawls’s *A Theory of Justice*¹³, like her reading of Kant (unsurprisingly, as Rawls himself borrows extensively from Kant), represents political conflict as something that can, in principle, be resolved. She writes that

“Rawls imagines a political culture whose fundamental conflicts (between liberty and equality) are ‘settled’; a civic culture that engenders no resistances to itself; a practice of punishment that sparks no remorse [...] If we have trouble envisioning this ideal, it is because our intuitions are contradictory and in need of rational systematization or

¹³ It is worth noting that Honig was unable to examine the reworked conception of a just society elaborated in *Political Liberalism*, as the latter was published around the same time as Honig’s *Displacement*.

because our perspective is skewed, corrupted by the unjust institutions in which we live.” ([1993] 2023, 127)

Importantly, the second part of this quotation highlights rational reasoning as the source of Honig’s central dissatisfaction with Rawls: she argues that his universalistic conception of rationality seeks to resolve all issues neatly, leaving no space for remainders: “This is the task Rawls assigns to politics and to political theory: to isolate the right vantage point, to establish the right setting, to facilitate the identification of these right resolutions, to dissolve the remainders of politics rather than engage them” ([1993] 2023, 127).

Honig critiques Rawls’s assumption that all it takes is the construction of well-ordered institutions, which will naturally produce well-ordered individuals who exercise the self-restraint and self-concealment that (Kantian) rationality demands. That is, Honig argues that Rawls assumes a rational institutional framework will ensure individuals internalise its norms without friction. However, Honig contends that a world without resistance is impossible, because depoliticising individual subjectivity (through the expectation of individual responsibility qua rationality) will not resolve the resulting dissonance ([1993] 2023, 127–128). The most obvious reason is that the individual cannot simply transcend the institutional abstractions designed to serve not a single unit but the social body as a whole.

Honig’s challenge to Rawls’s conception of the rational subject underpins her broader critique of his treatment of pluralism. She points out that Rawls, by claiming that foundational pluralism, i.e., divergent comprehensive visions of the good life, cannot be reconciled, concludes that it must be excluded from the political sphere. This assumption, however, rests on the view that citizens accept their foundational commitments as peacefully coexisting lifestyles. Moreover, Honig contends that Rawls depoliticises pluralism, a move she sees as harmful in “an age of social homogenization”, where “spaces of plurality and difference are often closed by the sedimentation or naturalization of a dominant culture’s form of life” ([1993] 2023, 130). For Honig, pluralism necessitates political engagement – a notion that, she argues, Rawls not only overlooks but undermines, since its depoliticisation enables domination to persist by normalising certain social conventions (for instance, those embedded within gendered structures) while marginalising others (ibid.).

Thus, the problem of Kantian-Rawlsian rationality as the means by which resolutions are achieved is that it becomes the source of authority, while its limitations are merely obscured (see, e.g., [Brubaker \[1984\] 2010](#)) – most notably, an unwavering need for closure (as in a particular conception of justice) and the ensuing rationalisation, that is, post hoc reasoning, which serves to justify

closure-driven decisions. This, on Honig's account, leads citizens to become "relatively passive consumers of the state's goods and services, impervious to the possibility that their survival as a democracy might require them to engage and resist" ([1993] 2023, 129).

There is one further important concept of Rawls's that Honig addresses – the sense of justice ([1993] 2023, 140–141). This is a notion Rawls assumes everyone possesses, though individuals may disagree about its content. More precisely, Rawls speaks of a public sense of justice, which makes "secure association together possible" ([1971] 1999, 5). For Rawls, this sense is both a sentiment (*ibid.*, 44) and a skill cultivated through rational deliberation and supported by fair institutional design in the exercise of judgement; it is this very capacity that the principles of justice as fairness seek to enable and affirm (*ibid.*, 41). The crucial point, however, is that everyone is presumed to possess it.

Honig, by contrast, argues that the expectation that moral sensibility will always be affirmed is misguided. She observes: "On Rawls's account, talents and attributes are randomly and inequitably distributed among persons, whereas the sense of justice is assumed to be equally available to everyone" ([1993] 2023, 141). The problem, as Honig sees it, is that if the sense of justice is a human attribute, and no human attributes are equally distributed, it would be unreasonable to expect moral sensibility to be uniformly shared.

The focus on moral sensibility reflects Honig's broader critique of Rawls, namely that this sensibility presupposes uniform agreement rather than conflict, thereby rendering it blind to values that fall outside the liberal consensus. Rawls assumes that "a well-ordered society tends to eliminate or at least to control men's inclinations to injustice" ([1971] 1999, 215), thereby conveying the hope that all that is required to eradicate the unjust element is the proper construction of fair principles of justice. Those who exist beyond the principles of liberal justice – that is, those lacking the sense of justice – are deemed by Rawls "irrational" and "extrasystemic" ([1993] 2023, 143), or even "sociopathic" (*ibid.*, 142). Honig aptly characterises these individuals as "irresponsible rogues and idiosyncratic misfits" (*ibid.*, 149). Consequently, on Honig's view, although Rawls's theory purports to accommodate pluralism, it ultimately recognises only pluralism within the bounds of liberalism, and not beyond it.

Honig's view of Sandel

Honig's ([1993] 2023, 162–199) reading of Sandel highlights his similarity to Rawls in their attempts to envisage a determinate social consensus. The difference between Sandel and Rawls lies in their respective ontologies of subjectivity, since while Rawls posits a one-dimensional, autonomous, and self-disciplining

individual, Sandel conceives of the self as having at least two additional dimensions, namely the intrasubjective and the intersubjective. The latter refers to the self as constituted by the community (*ibid.*, 163), while the former denotes the inner, and at times conflicting, plurality of identities within the individual. In other words, Sandel advances a pluralist conception of the self, in contrast to Rawls's singular account. However, Honig contends that although Sandel claims to reject Rawls's abstract self, he ultimately reproduces the same abstraction in a different guise, presuming that the individual can impose order upon their multiple selves through rational interpretation, thereby constructing a coherent and unified agency. She writes:

“Sandel replicates the Rawlsian move: he domesticates his intrasubjective conception of the self by way of a practice of self-interpretation that is strikingly similar to Rawlsian practices of self-ordering... the communitarian self is multiply constituted, but its multiplicity can be ordered and bounded, even univocally, in the recast original position; this self is not so radically multiple that it resists or exceeds, or somehow escapes the self's efforts to order itself (under the right conditions) into a coherent and unified agency.” (Honig [1993] 2023, 163).

Regarding Sandel's conception of subjectivity, Honig critiques his account of intersubjectivity for adopting an uncritical stance towards the nature of identity. She argues that Sandel, like Rawls, presupposes a subject that ultimately appears whole and unfractured – one devoid of the internal tensions that might disrupt its social essence in relation to community. As Honig observes, “Sandel's subjects are constituted and situated, but not produced; encumbrance is a source of meaning and self-understanding, but never of mystification or self-alienation, except when community is subverted by something outside itself” ([1993] 2023, 164). In other words, Sandel envisages a subject that simply inherits moral orientations and thereby attains stability, rather than one whose identity is constructed through contestation, conflict, or encounters with otherness. Honig aptly illustrates the tensions inherent in the construction of identity through otherness: “blackness plays a part in the constitution of whiteness, masculinity in femininity, master in slave, terrorist in state, and overman in herd” (*ibid.*, 171).

Sandel's conception of intersubjectivity is further complicated by intrasubjectivity, which arises from the notion that an individual is never a member of only one community but of many. As Sandel observes, “each of us moves in an indefinite number of communities, some more inclusive than others, each making different claims on our allegiance” (Sandel [1982] 1998, 146). It follows, for Sandel, that “the constituted self is constituted along so many

dimensions of subjectivity that she cannot know herself” (Honig [1993] 2023, 175, original emphasis), and consequently, “cannot heal herself without help [of the community]” (ibid.). Ultimately, Honig critiques Sandel for treating these multiple identities as a problem to be resolved, assuming that responsible agency requires a coherent, unified self. In doing so, he overlooks the possibility that inner conflict and multiplicity within the self can, in fact, serve as sources of empowerment (Honig [1993] 2023, 175).

As noted previously, Honig emphasises that both Rawls and Sandel seek resolution in the face of disagreement. But rather than invoking the Rawlsian notion of an imagined original position, Sandel envisions a form of ‘political friendship’ as a means to this end, while, as with Rawls, Sandel’s conception of politics entails “deliberation under certain ideal circumstances, ideal because they presuppose and secure the isolation, acknowledgment, and affirmation (but really the production and maintenance) of a morally relevant description of the self” ([1993] 2023, 176). Like Rawls’s original position, Sandel’s politics of friendship presupposes a like-minded social body willing to enter into the process of resolution in the first place. Yet, in Sandel’s case, it remains unclear how, within a context of intrasubjectivity – where different identities (for example, Orthodox Jews living in a modernised, gender-progressive society) may come into conflict – such tensions are to be addressed, unless they are excluded from deliberation altogether. Thus, Honig ultimately contends that Sandel “ignores the fact that in the case of the intrasubjective conception there is no clear and unproblematic answer to the question, who are my real friends?” (ibid., 178), which then “compels politics as friendship to renege on its promises of openness and deliver only on its promise of closure” (ibid.).

4.2.3 Honig’s agonism

To unpack Honig’s conception of agonism, let us begin at the end. While Honig is intensely opposed to fixing determinate resolutions in politics and instead advocates the continuous contestation of dominant structures, it does not follow that, for her, politics should concern itself solely with *virtù*. Rather, she acknowledges that politics requires both order and chaos, justice and disruptive action, and that it is not a matter of balancing one against the other, but of working within the persistent tension between them:

“Politics consists of practices of settlement and unsettlement, of disruption and administration, of extraordinary events or foundings and mundane maintenances. It consists of the forces that decide undecidabilities and of those that resist those decisions at the same time. To reduce politics to only one side of each of these operations, to

depoliticize the opposite side by identifying it with nature (as Arendt does) or with an outside agitation (as Rawls does), is to displace politics, to deny the effects of power in some of life's arenas for the sake of the perceived goods that power stabilizes under the guise of knowledge, respect, rationality, cognition, nature, or the public-private distinction itself" (Honig [1993] 2023, 205).

In an interview with Alyson Cole, Honig outlines how she understands agonistic opposition and disruption as capable of yielding positive outcomes:

"It is a key commitment of agonism that we make our opponents' positions the best they can be (the opposite of a straw man) and then see how engaging with their positions shapes our own. Rising to their challenge means we are in a relation of indebtedness to them." (Honig and Cole 2021, 345)

I propose to characterise this process as *adversarial sparring*. Engaging with an 'alien' position, and having to state one's own clearly enough for an alien interlocutor to grasp it, forces adversaries to improve their understanding of their own position and its foundations. Adversaries therefore seek not only to understand the other's position, or to find fault with it, but to deepen their grasp of their own. In this kind of dialogue (or, more precisely, polylogue), an opponent's objections become useful, since they reveal hidden assumptions and inconsistencies, and show how far one's commitments can be challenged before they dissolve.

Agonism, on this view, is a relational approach to politics. Adversaries are co-constitutive, helping each other clarify what they believe, why they hold it, and where it may need revision. Accordingly, political disagreement is not simply a contest between opposing views, but a shared process of questioning and learning from one another. When adversaries approach disagreement in this way, it becomes valuable not only for gaining knowledge but also for motivating people to develop their own views, so that disagreement is no longer viewed as threatening one's identity, but as a way of developing it. This, I suggest, is part of what Honig means when she describes politics as joyful ([1993] 2023, ix). The joy does not come from reaching agreement or defeating an opponent, but from recognising that an adversary can contribute to one's own political and moral development. We can view this disposition as one of *sparring partners* who challenge each other without trying to destroy one another. Such sparring enables one to gain a clearer understanding of one's own commitments and to become better able to defend or revise them. As Honig writes:

“This kind of joy comes from breaking out of bad psychological or political patterns that seek out control, sovereignty, mastery or subservience. Joy comes from the surprise of finding commonality of purpose in – or in spite of – contexts of difference or division.” (Honig and Cole 2021, 345)

That said, Honig is also self-critical, acknowledging that “the virtù perspective is vulnerable to the same charges [as moralist theories]: it is itself an instrument of consolidation” ([1993] 2023, 201). The idea is that although virtù is the ability to disrupt politics, it can only work when there are stable structures in place. At the same time, when these structures become too stable, they create the tensions that virtù responds to. As Honig puts it, “the lever of destabilization virtù provides is enabled by this stabilizing series of oppositions, a series that is itself vulnerable to subversion, to the disruptions of virtù” (ibid., 200). Virtù, in other words, needs the very stability it disrupts. However, at the same time, confronted with a weak point in an existing order, virtù does not simply destabilise it but reorganises it, since contesting stability first requires the contestor to establish categories of its own (ibid., 201–202) – which makes virtù, at the moment of disruption, an instrument of consolidation.

To conclude, Honig’s central argument is that while “the lesson of the contest of virtue and virtù is that politics never gets things right, over, and done with” ([1993] 2023, 204), the crucial task lies not merely in attempting to resolve ambiguities and struggles, nor solely in interrogating those struggles (as, for instance, a Geussian realist might urge us to do), but also in interrogating stability itself. This entails questioning the very existence of stability in order to uncover the underlying factors that sustain it – insights that may allow us to identify and exploit points of pressure, including seemingly invisible ones, inherent within that stability, thereby generating productive disruption and challenging our existing assumptions.

4.3 WILLIAM CONNOLLY’S AGONISTIC RESPECT

4.3.1 Connolly’s descriptive and explanatory account of antagonism

Among the theorists of political agonism, William Connolly’s account engages most directly with ethical questions. At the core of Connolly’s approach are the ethical values of agonistic respect and critical responsiveness, accompanied by the related notions of genealogy and the importance of ethical work upon the self (Wenman 2013, 100). These concepts are grounded in his fundamental

assumption that human suffering and conflict are “rooted in the human condition itself” (Connolly [1991] 2002, 91) and not merely in social structures:

“Humans resent the transiency, suffering, and uncertainty of redemption that mark the human condition. We suffer from the problem of our meaning, and we demand that meaning be given to existential suffering. So when the idea of a purpose in existence residing in nature or a god loses its credibility, the insistence that we are rational, responsible agents comes into its own. For if these previous sources of responsibility are dead, some new agency must be created. We give meaning to existential suffering, then, by holding ourselves responsible for it.” ([1991] 2002, 79)

Connolly, drawing on Nietzsche’s concept of *slave morality* ([1887] 2007, §§1.7, I.10), then moves from considering suffering a foundational state of affairs to the idea that the reconciliation of this state in the social sphere cannot materialise as a revolt against the existing dominant power, purely because “social powerlessness makes actual revenge untenable” (Connolly [1991] 2002, 79). This powerlessness then gives rise to individual ethical responsibility and self-discipline, wherein one interprets and shapes one’s own suffering as a source of personal meaning rather than seeking to overturn external powers (*ibid.*). Suffering is ever-present, and so, therefore, is doubt about the truth. Individual responsibility follows when the “autonomous individual – denied recourse to customary or rational foundations – confronts every moment as a potential ethical dilemma” (Wenman 2013, 106).

A person who sees themselves as responsible for their own suffering is likely to regard others as equally responsible for theirs. Connolly writes: “The modern individual, in short, contains resentment against the human condition in its own identity, and this comes out most clearly in the intensity of the resentment it expresses against any others who deviate significantly from that identity” (Connolly [1991] 2002, 80). Put differently, human beings resent the indeterminacy of their own identity – while the indeterminacy is fundamental to the human condition – and, in struggling with this inability to secure identity, they project resentment onto others who differ from them. Consequently, social conflict becomes rooted in perceived deviations from a presumed desirable identity. By extension, this fosters hostility towards those who interpret suffering as the product of abstract social structures, since the source of suffering is perceived not to lie within these structures but within the individual. As Wenman observes: “Connolly tends to explain the manifestations of socio-economic injustice as largely a consequence of existential anxiety, and so he looks

for solutions in personal ethics rather than in the possibility of a transformation in the basic structures or processes of society” (2013, 101).

Connolly ([1993] 2002) draws extensively on Nietzsche to explain the emergence of the contemporary dominant moral system in the West, and on Augustine to delineate its limitations, developing an account of how particular conceptions of morality shape individual resentment and, in turn, engender identity-based agonism within the social sphere. In short, the search for a ‘true’ moral self (in Augustinian terms) produces not merely a diversity of identities but also animosity. The tension generated by seeking ‘trueness’ in the self is projected onto others, who are then regarded as heretics because they do not share the same ‘trueness’ (understood as an intrinsic moral order, i.e., not requiring further justification). Since moral truth must, by definition, exist, these others are assumed to have some truth; yet, as it is not the correct one, they are cast as enemies. Morality is thus deeply entangled with identity – indeed, it underpins it. Those who do not share the truth are treated as heretics (Connolly [1991] 2002, 3).

Moreover, Connolly suggests that this division fulfils a profound psychological need, namely the displacement of blame from oneself onto others for one’s suffering, that is, failure to realise one’s true identity and do as one desires. Constructing the ‘liable other’ provides the sufferer with a rationale for animosity – as he notes: “*Human beings resent the transiency and suffering which define the human condition. This condition can be tolerated best if humans can find some agent who is responsible for suffering, an agent who can become the repository of resentment*” (Connolly 1994, 153, *original emphasis*). In society, this translates into dominant groups maintaining a constructed ‘normal’ identity and disciplining citizens into conformity by mobilising the authority of state institutions to designate a range of both external and internal identities as “foreign”, particularly those regarded as disruptive to prevailing social norms and values ([1991] 2002, 207).

4.3.2 Connolly’s normative account of agonism

Thus far, we have examined Connolly’s account of the natural condition of human beings and society, which he characterises as intrinsically antagonistic. This antagonism arises from a fundamentalist conception of self and identity, grounded in the presumed truth-aptness of a fixed axiological framework. It consequently generates resentment (or resentment in the Nietzschean sense) towards identity-others, who are held responsible for the failure to attain a ‘true’ identity. We may now turn to Connolly’s normative perspective, considering how he envisions a transition from resentment – understood as a sense of grievance over unrealised social or political possibilities ([1991] 2002, 191) – to

agonism through pluralisation, achieved by means of “agonistic respect” and “critical responsiveness” (1995, xx).

Connolly’s normative point of departure is a critique of liberalism’s toleration of difference and pluralism, which he regards as flawed because it ultimately sustains the status quo. This objection is not unique to Connolly but is shared by other critically oriented approaches, particularly strands of feminist theory (e.g., Abbey 2015), which argue that (Rawlsian) liberalism’s emphasis on public justification – making positions acceptable within prevailing norms – constrains, rather than enables, the emergence of new or previously unacknowledged identities and relations. In Connolly’s words:

“[...] liberalism is attentive to ‘difference’ as already defined, heterodox identities in an existing network of social relations, but tone-deaf to ‘différance’ as that which resides within an existing network but has not yet received stable definition within it. As a result of this thinness in the appreciation of difference, dominant forms of liberalism remain unattuned to the crucial role new enactments or performances play, first in generating new claims to identity, and second in retroactively crystallizing violence in previous patterns of being.” (1994, 178)

In other words, liberalism recognises already established social distinctions (e.g., race, gender, sexuality, religion), while remaining deaf to *différance* (in the Derridean sense) ([1991] 2002, 13–15, 49–50), where identities are not yet fixed but are still in the process of being formed. The implication here seems to be that liberalism affirms the status quo, including the crystallisation of existing oppression, by implicitly suppressing transformational processes. In doing so, it keeps such processes unexpressed, thereby allowing antagonism to brew beneath the surface.

Connolly’s conception of agonism is intended to address forms of difference that have not yet become manifest or been acknowledged. He contrasts his conception of agonism with the two traditional lenses through which social cohesion has typically been understood, namely individualism, which privileges rational collaboration while relegating differences of intrinsic identity to the private sphere, and communitarianism, which rests upon strong cultural bonds sustained by an exclusive morality (1995, xx). Connolly argues that the “possibilities for democratic political action are enhanced when diverse, interwoven constituencies resist becoming frozen into contending claims to intrinsic identity or exclusive morality” (ibid.).

Connolly’s substantive normative approach to addressing antagonistic resentment in society is that of pluralisation (1995), which centres on two

questions: “asking how to foster an ethos of reciprocal forbearance and responsiveness in relations between interdependent, contending identities [and] how to pluralize lines of territorialization so no single line becomes overcoded” (ibid., xxiii). This means that people with different and conflicting identities should learn to tolerate and respect one another because they depend on each other, and that society should not become fixed around one dominant way of organising or living but instead leave space for a variety of ways of living and belonging.

4.3.3 Critical responsiveness and justice

Connolly includes ethics in his account as a requirement for pluralisation. He argues that agonism needs an ethical outlook that is neither a modernist acceptance of fixed moral foundations nor a postmodernist rejection of them. Specifically, he refers to “contractual morality, command morality, or teleological morality [of which] the modernists find these options to suffice; the posties find all three perspectives to harbour too many violences and exclusions” (1995, xxiii). On this basis, Connolly distances himself from Rawlsian, Kantian, and consequentialist reasoning, broadly agreeing with postmodernists that such approaches rest upon rigid metaphysical frameworks of ultimate purpose and thereby tend towards exclusion and domination. Instead, he proposes a fourth possibility: an *ethic of virtue*, but one that is detached from the teleological metaphysics that posit an “intrinsic purpose” (ibid.). In Connolly’s own words, virtue:

“pursues an ethic of cultivation rather than a morality of contract or command; it judges the ethos it cultivates to exceed any fixed code of morality; and it cultivates critical responsiveness to difference in ways that disturb traditional virtues of community and the normal individual. It does not present itself as the single universal to which other ethical traditions must bow. Rather, it provides a prod and counterpoint to them, pressing them to rethink the ethics of engagement and, crucially, to rework their relations to the diversity of ethical sources that mark a pluralistic culture. Such a post-Nietzschean ethic resists oligopolistic control over the currency of morality while affirming the indispensability of ethics.” (1995, xxiv).

For Connolly, the primary distinction between an ethic of virtue and what he terms the ‘traditional’ moral codes lies in their orientation towards difference. The former emphasises critical responsiveness to difference, refusing to accept the existing moral order (as the basis of identity) as given and subjecting it to critical challenge, whereas the latter seeks to establish a fixed and universal

moral code. The object of Connolly's reflection is the difference of identity, which, according to him, cannot be resolved through universalised principles but instead requires the cultivation of an ethos of engagement. It is the virtue of responsiveness that grounds his account of a positive programme for engaging with potentially – or more precisely, inevitably – conflictual differences.

He identifies three features of *critical responsiveness* as crucial for working towards an agonistic pluralisation, namely anticipation, critical engagement, and self-revision (1995, 184). *Anticipation* refers to a responsiveness to difference before that difference has been fully formed or recognised as a stable identity. Groups that at some point become consolidated into acknowledged identities (e.g., LGBTQ+) typically begin as movements struggling for recognition. Agonism reads these movements while they are still taking shape, rather than after they have become established, guided by the assumption that identities develop through political struggle. *Critical engagement* does not entail unconditional acceptance of every new movement but instead involves evaluating whether such movements genuinely pursue pluralism or whether they risk becoming exclusionary or dominant. *Self-revision* (in other words, reflexivity) designates a mode of engagement with emerging identities that requires conceptual transformation on the part of those occupying established positions. If new differences are to be recognised, dominant groups must alter their own self-conceptions, relinquishing assumptions that previously depended upon exclusion. This entails a de-essentialisation of existing dominant identities, revealing that categories such as 'normal' are historically constructed in opposition to those designated as 'the other' in order to justify one's dominance (see also Foucault 1978).

The importance of critical responsiveness in addressing resentment-driven antagonism is demonstrated in Connolly's account of political justice (1995, 184–186). Prior to recognition, rights-based conceptions of justice (as in liberal contractualism) often reinforce existing injustices, since differences that are not yet recognised are simply excluded. In this sense, the modern approach to morality, that is, "contractual morality, command morality, or teleological morality" discussed earlier, reproduces practices of exclusion (ibid., xxiii). Liberal rights-based reasoning, Connolly argues, is always one step behind recognition. For example, in his discussion of universal human rights (ibid., 185–186), he contends that grounding rights in a single moral condition – human dignity – makes them function as a universal closure, thereby obstructing further pluralisation. Because notions such as 'dignity' or 'equality' are defined within prevailing contexts, their scope is historically limited. For instance, homosexuality was not encompassed within human dignity or equality at the time the Universal Declaration of Human Rights was drafted. It is therefore

political struggle, rather than universal morality, that secures the inclusion of new identities within the universal.

With this, Connolly warns against liberal narratives of inevitable moral progress (such as relinquishing discrimination based on sexuality). Interpreting such histories as Hegelian dialectics that guide us towards inevitable progress obscures the contingent and precarious struggles that made them possible, fostering self-congratulation over past victories while overlooking the continued exclusion of marginalised groups, such as those seeking the right to die with grace (1995, 186–187).

Thus, justice understood in the rights-based sense of moral universalism cannot transcend the existing moral order, as it denies the political struggles through which that order has been constituted. Hence, critical responsiveness is needed. As Connolly observes: “The ethos of critical responsiveness is more fundamental than justice partly because its very shape calls into question the quest to find a solid ground of morality or justice” (1995, 187). Unlike liberal justice, critical responsiveness acknowledges that identities are never fixed but always relational and historically contingent: “For it is this fugitive, always underdeveloped, experience of contingency in what you are that taps into the ethical capacity to respond to injuries shuffled under justice by the practice of justice” (ibid., 188).

This does not imply, however, that critical responsiveness alone is sufficient, for it merely acknowledges difference without providing practical application. For this, justice remains necessary. Moreover, critical responsiveness and justice are interdependent, but also in tension. Connolly observes: “Justice and critical responsiveness are thus bound together in a relation of dissonant interdependence. Reforms in justice depend upon a more general ethos of critical responsiveness to new social movements, and the practice of justice invokes relations of agonistic respect between contending identities already on its register” (1995, 187). In other words, the norms of justice cannot, on their own, accommodate previously unacknowledged claims; they depend upon critical responsiveness and its openness to new identities. Conversely, justice presupposes negotiation between different identities, requiring what Connolly calls *agonistic respect* – a mode of engagement in which entities that disagree, but are already recognised and possess some social power, contest one another while acknowledging each other’s legitimacy (ibid., 235). In short, critical responsiveness brings recognition to the table, and justice brings respect.

To summarise, the moral foundation of Connolly’s agonism is the call to replace moral fundamentalism, which seeks objective moral truth, with an ethical reasoning that takes contingency – not fixed truths – as its fundamental premise: “We [the agonists] think an ethical sensibility is to be cultivated because there is

no compulsory basis for ethics” (1995, 235 fn. 36). In an agonistic society, this presumes that contestability is reciprocal – that is, all sides acknowledge that they compete with each other, but no moral framework is absolutely final or beyond question. It is precisely this reciprocity that makes agonism possible, creating space for genuine, ethically driven interaction (ibid., 188).

4.4 CHANTAL MOUFFE’S AGONISTIC PLURALISM

Mouffe’s conception of agonism emerges from the post-structuralist approach she developed with Ernesto Laclau within the framework of what they term radical democracy (Laclau and Mouffe [1985] 2001). This framework conceives of politics as the continuous construction of meaning through hegemonic practices and understands ethics as shaped by historically and culturally contingent discourses. Within this context, agonism denotes a political stance in which actors who hold fundamentally opposed and irreconcilable moral worldviews are regarded not as enemies to be destroyed, but as legitimate adversaries with whom engagement is both possible and necessary (Mouffe 1999).

Building upon this theoretical foundation, Mouffe’s account of agonism is primarily concerned with the design of democratic institutions capable of managing irreconcilable moral conflicts without erasing disagreement. For Mouffe, democracy must insist on institutionalising a space for ongoing contestation among adversaries who, while sharing basic democratic commitments of liberty and equality, fundamentally disagree on their interpretation. It is this divergence of interpretation that Mouffe identifies as a genuine political disagreement – the core of political action. She critiques mainstream liberal theories, particularly those of Rawls and Habermas, for attempting to eliminate or domesticate antagonism by constructing a rational consensus that effectively forecloses the space for such disagreement (1999). While Rawls treats contingency instrumentally – as a dimension from which reasons for consensus emerge (see chapter 2) – Mouffe considers it a flame that breathes life into democracy, as it sustains the passion for engaging in political participation (2000, 104).

In this respect, Mouffe’s work engages less with the conceptual and ontological foundations of agonism – as Connolly and Honig do – and focuses more on the practical organisation of political contestation and the ways in which pluralism may be channelled constructively within democratic systems.

4.4.1 Radical democracy and its open-endedness

Before examining Mouffe's conception of agonism, it is useful to outline briefly her and Laclau's broader project of radical democracy, within which Mouffe's later work is situated. This project constitutes a normative and emancipatory endeavour aimed at uncovering the dynamics of what they describe as the "game of hegemony", in which "the rules and the players are never fully explicit" (Laclau and Mouffe [1985] 2001, 193). Their approach is at once critical and prescriptive, yet distinctly post-foundational, rejecting any appeal to fixed or universal political grounds.

Radical democracy advances a sustained critique of liberal political theory – particularly the work of Rawls and Habermas (Mouffe 1999; 2005) – for presupposing the rationality of consensus and the possibility of comprehensive agreement. It exposes what Laclau and Mouffe regard as a central flaw in liberal democratic thought, namely the assumption that dismantling hierarchies on the basis of moral equality can in itself secure social and material equality. Moreover, they argue that it is incoherent to affirm a commitment to value pluralism while simultaneously maintaining the ideal of equality within a framework of popular sovereignty (Mouffe 2013, 123–124), but they also reject class-based Marxist theories for their essentialist reduction of political identity to structural or economic categories (Laclau and Mouffe [1985] 2001).

The anti-foundationalism of radical democracy entails a rejection of the claim that political or moral life can be grounded in objective or universal principles. Instead, it seeks to reveal the contingent and constructed nature of social order (Selg 2012, 89), challenging the appearance of consensus and equality as natural or self-sustaining. This perspective also repudiates any form of politics grounded in epistemic privilege – what Laclau and Mouffe describe as "a form of politics which is founded upon ... [an] implicit assumption of a privileged point of access to the truth, which can be reached only by a limited number of subjects" ([1985] 2001, 191–192).

From this perspective, social order is never final and fixed, but instead the temporary result of hegemonic practices through which particular interests are presented as universal within ongoing political struggles over meaning. The task of political theory, therefore, is not to uncover foundational truths (as, on this view, they do not exist), but to critically examine dominant ways of thinking and develop alternatives that broaden the boundaries of inclusion. Radical democracy therefore rejects fixed conceptions of the good, seeing them not as pre-political truths but as products of hegemonic construction. Consequently, democracy is considered truly democratic only if it remains permanently open to renegotiation. This position resonates with a Wittgensteinian insight: "there is

no such thing as the ‘application of a rule’ – the instance of application becomes part of the rule itself” (Laclau and Mouffe [1985] 2001, ix).

Radical democracy is described as ‘radical’ because it seeks to go to the root (from the Latin *radix*) of democratic principles, aiming to extend and deepen them into new areas of social life, rather than rejecting them outright, as, for example, a Marxist revolutionary approach advocates. Drawing on Laclau and Mouffe’s theory of hegemony ([1985] 2001), it aims to transform the prevailing liberal order by articulating diverse democratic demands – such as those related to gender, race, or labour – into a shared political project. In this way, radical democracy engages in an immanent critique of liberalism, as it recognises the structural limitations of liberal democracy while seeking to re-signify its core values of liberty and equality in more inclusive and pluralistic terms (Mouffe 2013, 133–134). Creating new hegemonies involves bringing different struggles together through a discursively constructed chain of equivalence, forming a collective political subject capable of challenging existing power relations (Laclau and Mouffe [1985] 2001).

The framework also reflects a post-Marxist commitment to the idea that political differences and pluralism cannot be reduced to a single underlying conflict. It rejects the notion of a single, overarching antagonism – such as class – and instead emphasises a multiplicity of intersecting struggles (e.g., class, race, gender) that cannot be reduced to a single axis of conflict and resist resolution through consensus (Laclau and Mouffe [1985] 2001). These struggles are rooted in socially constructed group affiliations rather than essentialist identities, with meaning and justification emerging from the social practices that generate and sustain these hierarchies. As such, while Laclau and Mouffe’s discourse theory is typically regarded as emblematic of post-structuralist thought – emphasising difference, contingency, and the impossibility of closure – it nonetheless retains structuralist features, notably its claim that the social world is organised around group-based identities and antagonisms.

A central question for radical democracy is whether meaningful political contestation can take place in the absence of any shared common ground. Although the foundational pluralism of radical democracy emphasises divisions rooted in substantive interests, it nonetheless assumes the existence of a minimal level of shared agreement. As Mouffe puts it, “What exists between adversaries is, so to speak, a conflictual consensus – they agree about the ethico-political principles which organise their political association but disagree about the interpretation of these principles” (2013, 139). Yet Mouffe does not clearly specify what these “ethico-political principles” entail, referring only to liberty and equality, and doing so in rather indeterminate terms (Erman 2009, 1044). She

merely acknowledges that any substantive political agreement – understood more broadly as order or stability – remains provisional (2013, 2). Nonetheless, even provisional forms of agreement presuppose some degree of enduring consensus – namely, a shared commitment to the very possibility of reaching agreement. Erman further contends that conflict presupposes not a “minimally shared normative consensus”, but a shared understanding of “what is at stake” (2009, 1047), thereby calling into question the coherence of Mouffe’s proposed transition from antagonism to agonism. The following subsection examines and critically evaluates this issue in greater detail.

4.4.2 Conflict

A central question in the study of radical democracy concerns the extent to which political conflict can be meaningful. One of the most distinctive features of radical democracy underpinning Mouffe’s agonism is its reconceptualisation of political conflict. Drawing on Carl Schmitt’s notion of the political as rooted in the friend-enemy distinction, Laclau and Mouffe argue that the presence of conflict that cannot be fully resolved is constitutive of the political (Beckstein 2011). To understand this claim, it is necessary first to unpack the distinction Mouffe draws between the political and politics.

Mouffe conceives of the political as the “*dimension of antagonism that is inherent in human relations, antagonism that can take many forms and emerge in different types of social relations*” (2000, 101, emphasis added). By contrast, politics refers to the ensemble of practices, discourses, and institutions that seek to establish order and organise human coexistence under conditions that are always potentially conflictual (ibid.). Politics, in this sense, functions by “domesticating hostility and ... trying to defuse the potential antagonism that exists in human relations” (ibid.). The antagonistic world of the political is therefore characterised by persistent indeterminacy, in which power occupies an “empty space” (ibid., 2) and where “a final guarantee, a definite legitimisation” is impossible (ibid.) – that is, a world in which no single overarching authority or ideal can claim absolute legitimacy or dominion over others. In brief, the political is the realm of an inherent social state of conflict, while politics is about organising society by managing, but never fully eliminating, that conflict.

For Mouffe, social relations encompass two distinct types of disagreement: (a) oppositions grounded in foundational, “non-negotiable moral values” (1999, 756); and (b) oppositions concerning political, including ethico-political, principles (2013, 7–8).

(a) Conflicts over foundational moral values, Mouffe argues, do not belong to the realm of the political precisely because they lie beyond the scope of rational

deliberation. Yet they remain indispensable to the political process, as they provide the affective energy that animates political engagement. Passions, understood by Mouffe as forms of collective affect emerging from individual emotions (Zembylas 2021), are therefore a necessary – though not sufficient – force binding individuals to political projects. Disagreement arising from irreconcilable value pluralism should not be eliminated, for its resolution would extinguish the very passions that sustain democratic participation. Rather, such passions must be channelled and mobilised in the service of democratic design (2000, 103). As Mouffe cautions, “Too much emphasis on consensus and the refusal of confrontation lead to apathy and disaffection with political participation” (2000, 104).

(b) By contrast, conflicts concerning the organisation of common social life lie at the heart of liberal democracy. Such conflicts “cannot and should not be eradicated” (2013, 7), for the “confrontation of democratic political positions” is essential to democracy itself. Without this ongoing contestation, Mouffe warns, “there is always the danger that this democratic confrontation will be replaced by a confrontation between non-negotiable moral values or essentialist forms of identification” (ibid.), which inevitably leads to political deadlock.

When Mouffe speaks of *agonistic* conflict, she refers specifically to disagreement that remains within the (b) organisational bounds of rational negotiation – that is, conflict over political rather than moral issues. A key task of agonism, then, is to distinguish non-negotiable moral values from political ones and to exclude the former from the political sphere. This distinction is crucial, as Mouffe warns that introducing morality into politics risks shifting contestation onto a moral register. She argues that there is a “direct link between the weakening of the political frontier characteristic of the adversarial model and the ‘moralisation’ of politics” (2005, 75). When democratic confrontation no longer involves a shared commitment to the ethico-political values of liberty and equality – focused instead on disagreement over their interpretation (ibid., 121) – it becomes a struggle over what is morally ‘good’ or ‘evil’ (ibid., 75). Such moralised conflict is inherently irresolvable and produces deadlock, fostering polarisation in which opponents are viewed as enemies rather than adversaries. For example, when right-wing populism is framed as intrinsically ‘evil’ and liberalism as inherently ‘good’ (ibid., 73), both sides ultimately reject shared political values (ibid., 121). As Mouffe notes, “when opponents are defined not in political but in moral terms, they cannot be envisaged as an ‘adversary’ but only as an ‘enemy’. With the ‘evil them’ no agonistic debate is possible, they must be eradicated” (ibid., 76). By contrast, agonistic democracy rejects moral framing while acknowledging the existence of conflicting moral commitments, contesting their interpretation instead within a shared political framework (ibid., 121).

4.4.3 From antagonism to agonism

As Selg notes, Mouffe's concept of agonism is not a descriptive theory but one with a clear normative ambition (2012, 97), as it seeks to prescribe political arrangements that transform the default condition of antagonistic social relations into agonistic ones through political means. This involves envisioning institutions capable of structuring such agonistic interactions, while also assuming a general civic disposition that regards the idea of friendly enemies as a desirable political attitude. Such a disposition is essential for legitimising any institutional design aimed at resolving conflicting interpretations of foundational value commitments, because Mouffe contends that foundational moral commitments are not merely unlikely to be reconciled but irreconcilable in principle (Erman 2009, 1040); in other words, their substantive content lies beyond reconciliation.

Consequently, constructing institutional arrangements that enable the contestation of substantively irreconcilable truths becomes central to radical democracy. The overarching agonist goal is to cultivate a "conflictual consensus", in which citizens agree to disagree on core values yet remain committed to collectively working through their differences (Mouffe 2000, 103–104).

Let us now explore the crucial presupposition that, for Mouffe, renders the transition from antagonism to agonism possible – namely, the *presupposition of an agonistic disposition*. As previously noted, Mouffe has been critiqued for positing this disposition as antecedent to the political process – that is, for assuming that citizens are permanently disposed in this way. Selg (2012) refers to this as an "ethos of contingency": a foundational set of values or sensibilities that, descriptively, conceptualise the social world as shaped by contestable ideas influenced by hegemonic structures, and normatively, as requiring constant interrogation of prevailing truths. In this respect, such an ethos is neither latent, passive, nor reactive; rather, it is *proactive*, actively seeking to challenge dominant principles of political order.

In this sense, Mouffe's agonism functions as a civic ethos. Yet, because this disposition is not given *a priori* – for if it were, the project of radical democracy would be redundant – the question of how such an ethos might be diffused across the social field remains problematic. As Kapoor (2002, 472–473) observes, it is unclear how, for Mouffe, a civic predisposition towards agonistic engagement could emerge in contexts where repressive forces actively suppress its development. This difficulty arises because it lies within the interests of dominant power holders to maintain antagonistic social relations. For this, they construct oppositional narratives that shape value discourses in favour of the

hegemonic order. In brief, “political practice constructs the interests it represents” (Laclau and Mouffe [1985] 2001, 120).

Consequently, the very structures that Mouffe seeks to re-hegemonise also obstruct the conditions necessary for that transformation. Since her project is ultimately concerned with the redesign of democratic institutions, a critic would argue that agonism appears toothless if it cannot account for how political subjects are to become agonistic in the first place (Selg 2012, 96–97). This ambiguity arises, in part, from Mouffe’s treatment of foundational moral commitments as given starting points for political reasoning. By presupposing such commitments – such as equality or liberty – without explaining their origins or justification, her account risks leaving unclear how political subjects come to adopt an agonistic disposition in the first place. If the purpose of agonistic democracy is to contest divergent interpretations of these shared commitments – for instance, competing conceptions of justice – without seeking their reconciliation, then elucidating the processes and institutional conditions through which such contestation can occur should form a central concern of agonistic theorising.

Leaving this critique aside, let us turn to how Mouffe conceptualises antagonism in ontological terms, in order to clarify what a transformation to agonism would entail. When Laclau and Mouffe discuss the nature of antagonism, they draw on Lucio Colletti’s distinction between real opposition and contradiction (Laclau and Mouffe [1985] 2001, 123–125). They argue that antagonism refers neither to real, physical opposition (for example, two vehicles colliding – an event that is material rather than relational), nor to contradictory opposition, where A is not B simply because the totality of A differs from that of B. These forms of opposition are not truly antagonistic because they each represent determinate, self-contained states of being that exist independently of one another. They stand in opposition without any hostile disposition – that is, without the intent to negate or suppress the other’s totality (ibid., 124).

By contrast, antagonism reveals that identity and social relations are not constituted through completeness or self-sufficiency, but through conflict, dependency, and negation (Laclau and Mouffe [1985] 2001, 125). The antagonistic relation does not arise from the interaction of stable and pre-constituted identities; rather, it emerges from the impossibility of either identity ever achieving full realisation. As Laclau and Mouffe write:

“The presence of the ‘Other’ prevents me from being totally myself. The relation arises not from full totalities, but from the impossibility of their constitution. The presence of the Other is not a logical impossibility: it exists; so it is not a contradiction. But neither is it subsumable as a

positive differential moment in a causal chain, for in that case the relation would be given by what each force is and there would be no negation of this being. [...] Antagonism, as a witness of the impossibility of a final suture, is the 'experience' of the limit of the social." (Laclau and Mouffe [1985] 2001, 125)

Further, drawing on late Wittgenstein, Laclau and Mouffe argue that language – the medium through which antagonism is articulated – operates by establishing systems of difference, where meaning emerges only in relation to what something is not (for instance, democracy gains significance through its contrast with autocracy) ([1985] 2001, 125). Antagonism signifies the collapse of these very differences, for example when autocrats start calling themselves democrats, because it marks a breakdown within language itself and can consequently be expressed only metaphorically. In this sense, antagonism does not arise between social agents as such, but rather exists outside them, while nonetheless constituting the limits of society – specifically, revealing the impossibility of society ever becoming fully harmonious (i.e., interpretations will always vary). Antagonism is thus both external to society and foundational to its definition. It exposes that society can never be complete or self-identical; its identity remains perpetually open, contingent, and indeterminate. Paradoxically, then, antagonism is what defines the social order while remaining irreducible to it.

Antagonism thus reflects the paradox at the core of democratic logic, which requires drawing a boundary around the demos, between those included as 'friends' and those excluded as 'enemies'. As Mouffe observes, "what cannot be contestable in a liberal democracy is the idea that it is legitimate to establish limits to popular sovereignty in the name of liberty" (2000, 4). Put simply, democracy depends upon a continual, open-ended tension between inclusion and exclusion. To sustain democratic vitality, antagonism must not be eradicated, for it ensures that the democratic 'self' – its identity, interests, and desires – remains incomplete and open to contestation. This incompleteness arises not merely from the existence of a contradictory 'other', but from the interpretative and linguistic indeterminacy between 'self' and 'other', a relation that eludes the control of either.

This means that antagonism is not merely another category within society, such as class, gender, or race, nor is it simply another form of opposition. Rather, it is external to society in a structural sense – it reveals the impossibility of fully closing or suturing society into a final, unified whole (Mouffe 2000, 127). Two implications follow from this. First, society cannot be equated with an empirical totality; that is, the observation of individuals' behaviour within a given space does not yield an adequate representation of society. Second, antagonism cannot

be resolved through rational means, as it disrupts the notion of society as a rational and coherent totality. It is not a constitutive part of the system, as it would be within Hegelian or Marxist dialectics (ibid.), and therefore it is also not reducible to social ills, such as economic or social inequality or political exclusion (though it can be partially described in terms of them).

The concept of antagonism, moreover, calls for a closer examination of the distinction between two competing logics that structure political discourse: *equivalence* and *difference* (Laclau and Mouffe [1985] 2001, 127–134). In the logic of equivalence, features such as dress code, language, skin colour, or customs are mobilised to signify the presence of dominant power. As they explain, “[s]ince each of these contents is equivalent to the others in terms of their common differentiation [...], it loses its condition of differential moment, and acquires the floating character of an element.” (ibid., 127). These features are stripped of their particularistic meanings and rearticulated to serve a new function – namely, to construct a unified category of opposition or dominance. As Laclau and Mouffe put it, “equivalence creates a second meaning which, though parasitic on the first, subverts it: the differences cancel one other out insofar as they are used to express something identical underlying them all” (ibid.). In this way, equivalence simplifies the political space into binary oppositions, thereby generating antagonism. In contrast, the logic of difference organises and stabilises political demands into a more coherent – though never complete – social order (ibid., 128). Politics, then, continually oscillates between these two logics, as neither can fully prevail, underscoring the contingent nature of the social (ibid., 129).

In the most basic sense, “the logic of equivalence is a logic of the simplification of political space, while the logic of difference is a logic of its expansion and increasing complexity” (ibid., 129). These two logics may be understood as underpinning communitarian-conservative and social-liberal forms of politics, respectively, in that the former seeks social homogeneity, while the latter pursues social pluralism through the promotion of liberty.

Consequently, Laclau and Mouffe’s project rejects both approaches, on the grounds that each ultimately fails by seeking a form of determinacy – whether unificatory or pluralist. Instead, they advocate for a radical indeterminacy as the only means of achieving a truly democratic politics. This position stems from the understanding, within the framework of radical democracy, that antagonism reveals indeterminacy to be a pervasive condition – even within liberal emancipatory projects. Such projects often disregard internal antagonisms (e.g., within feminism) and focus solely on an external enemy (such as male privilege), thereby obscuring the complexities of internal conflict (ibid., 132).

Thus, we can infer from Laclau and Mouffe's ([1985] 2001) conceptualisation of antagonism that foundational disagreements remain irresolvable because the 'us vs. them' distinction is not merely an opposition between totalities, but rather signals the impossibility of fully constituted identities – even as political processes attempt to construct them. Antagonism does not represent a fixed structure or set of features within society; rather, it marks, through negation (i.e., what it is not), the limits of determination. Antagonism sustains democratic social relations by rendering irresolvability and permanent openness visible and operative. Thus, fundamental value – and moral – disagreements are not obstacles to radical democracy, but essential conditions for its continual renewal and contestation.

However, as we have already seen, antagonism is an inevitable, yet still undesirable, feature of political life. If neither the logic of equivalence nor the logic of difference alone can serve as an institutional arena for agonism, a crucial question arises: how can antagonism be transformed into agonism? As noted earlier, Mouffe's conception of agonism can be understood as a form of structural agonism, insofar as it focuses on how democratic institutions might cultivate spaces in which legitimate contestation over power and meaning can unfold. It is important to recall that, unlike Honig or Connolly, Mouffe assumes a minimal form of consensus where citizens must be united under the shared civic-liberal values of liberty and equality (2000, 102), thus presupposing that citizens share this disposition. Notwithstanding the earlier critique regarding how such values might emerge as common values if hegemonic structures work against their realisation, let us explore what this disposition entails.

To address this 'civic disposition', Mouffe introduces the concept of *symbolic space*. She argues that the purpose of politics is to transform antagonism into agonism – that is, to reconfigure 'friend-enemy' relations into relations between 'friendly enemies', in which the other is not to be eradicated (as the source of the problem) but recognised as a legitimate adversary: "The central task of democratic politics is to provide the institutions which will permit conflicts to take an agonistic form, where the opponents are not enemies but adversaries among whom exists a conflictual consensus" (2013, xii).

As Mouffe says, adversaries are friends and enemies at the same time: "friends because they share a common symbolic space but also enemies because they want to organise this common symbolic space in a different way." (2000, 13, emphasis added). An agonist friend is one who may contest the commitments of others while recognising their right to defend their ideas (ibid., 102).

Thus, what distinguishes antagonism from agonism is the "symbolic space" (2000, 13) shared by adversaries, but not by enemies. Within this symbolic space,

antagonism is not eliminated but rather reframed (ibid.). Yet what characterises this symbolic space, and what does it mean to organise it in a different way?

Although Mouffe does not explicitly define symbolic space, I interpret it as a domain in which political beliefs and ideas are rendered visible and intelligible – a cultural arena shaped not by shared values, but by a common system of representation. Such a system presupposes at least a minimal shared ethos – one that accommodates the agonistic disposition of friendly enemies. This entails engaging with opponents' ideas respectfully, which Mouffe describes as “the real meaning of liberal-democratic tolerance” (ibid., 102). For agonism to be viable, this ethos must be embedded within the fabric of society.

4.5 MORALITY AND POLITICAL AGONISM

As in political realism, agonism's conception of morality arises from its critique of the intrinsic character of the moral order – an order that attempts to fix particular conceptions of value and thereby privileges some over others, even in contexts where the reasons for such preferences are not unified. Because agonistic theorists emphasise different dimensions of agonistic politics, they also each highlight slightly different features of morality's nature. Nevertheless, they converge on the view that the drive to determine specific conceptions of the good – especially through the pursuit of rational consensus – ultimately extinguishes the vitality of moral and political life, a vitality that depends on ongoing contestation and pluralism. Let us now consider how the three agonist thinkers discussed here – Honig, Connolly, and Mouffe – conceptualise morality.

Honig emphasises that a stability-seeking moral order must continually manage – or even suppress – anything that challenges its claim to embody justice and thereby legitimise itself. This process operates not only through coercive mechanisms but also through everyday moral concepts such as respect, friendship, and responsibility ([1993] 2023, 3). In her critique of Kant, she suggests that respect is ultimately oriented less towards persons as such than towards the moral law or moral worth attributed to them, thereby helping produce compliant moral subjects rather than fully autonomous agents (ibid., 8).

Honig further contends that when morality is conceived as a categorical law grounded in reason, it inevitably becomes a constraining and illusion-producing force that “seals into place prevalent but unstable moral practices” (ibid.). In doing so, it constructs an ideal that is not even an objective one – if such a thing exists – but rather an ideal derived from, and therefore reinforcing, a particular historical conception of rationality. Kant's categorical imperative exemplifies this: “I ought never to act in such a way that I couldn't also will that the maxim on

which I act should be a universal law” (Kant [1785] 2008, G4:421). Viewed in this light, Kantian morality aligns with what Honig’s agonism rejects, namely the Nietzschean ‘slave morality’ of submission, meekness, and conformism (Nietzsche [1887] 2007, §§I.7, I.10).

However, Rawls should not simply be folded into Honig’s reading of Kant. As discussed in 2.2, although Rawls adopts Kantian rationality, he does not rely on Kantian duty in the same way, and his citizens’ moral powers function as capacities for fair cooperation rather than as obedience to a fixed moral law. Rawlsian citizens’ sense of justice functions as a source of motivation rather than as a substantive moral doctrine, and they are therefore not agents carrying out a pre-given moral law, but participants seeking fair terms of cooperation without presupposing any particular conception of justice.

Connolly’s conception of morality is most clearly articulated in two of his works: *Identity/Difference: Democratic Negotiations of Political Paradox* ([1991] 2002) and *The Augustinian Imperative: A Reflection on the Politics of Morality* ([1993] 2002). Morton Schoolman observes that Connolly, in a distinctly Foucauldian vein, reveals how the Augustinian conception of morality underlies the modern moral order, which, through its hegemonic governance of “our words and thoughts and deeds”, works to draw identity into a process of self-validation and dogmatisation (Schoolman [1993] 2002, xiv).

While Augustinian morality is fundamentally theological, Connolly ([1993] 2002) demonstrates how it endures through secularisation, showing that the idea of a single God as an authoritative standard of judgement has been adapted to shape and sustain the modern liberal subject. The Augustinian imperative emphasises purity, unity, and self-certainty through the subordination of desire, ambiguity, and difference (ibid., 46), while the “impurity” of imperfection is interpreted as a sign of error, corruption, or sin (ibid., 16), rendering the moral life a continual quest for redemption and unity (ibid., 46), through, for example, confession (ibid., 43–55), or in modern times, public acknowledgment. This strategy was adopted in modern moral and political thought through Enlightenment rationalism, with its underlying impulse to purify the contingent self and achieve coherence through reason, will, or universal law. Consequently, in Connolly’s view, the modern moral order remains Augustinian in form, albeit seeking salvation not in God but in rational self-affirmation, functioning as “obedience to transcendental command, and morality as attunement to an intrinsic design of things – by enunciating an alternative ethico-political spirituality” (ibid., xxviii).

Among the agonists considered in this study, Mouffe engages least with the conceptualisation of morality. Nevertheless, she acknowledges a distinction between what she calls the “moral universalistic” and “ethical pluralistic”

approaches (2000, 129). The former corresponds to (Kantian) “deliberative democrats”, with their emphasis on impartiality and rational consensus, while the latter – including thinkers such as Arendt and Nietzsche – defends the “unending pursuit of the recognition of the Other” (ibid.). Yet Mouffe argues that both perspectives miss the mark when it comes to politics, simply because their attention is directed not towards the political but towards the moral or ethical. As she states:

“What is missing in all of them ... is a proper reflection on the moment of ‘decision’ which characterises the field of politics ... They [decisions] entail an element of force and violence that can never be eliminated and cannot be adequately apprehended through the sole language of ethics or morality. We need a reflection of the political proper.” (2000, 130)

Thus, Mouffe rejects both moralists and their critics, who in her view fail to recognise hegemonic relations as the proper terrain of politics. She argues that any form of value-driven theorising diverts attention from what she considers the very core of social relations – namely, conflict. Like her fellow agonistic theorists, Mouffe traces the shortcomings of morality to the Enlightenment’s unwavering faith in rationality, which, she contends, produces an “idealized view of human sociability” (2000, 131). However, she also distances herself from both Honig and Connolly, noting that “it seems to me that their conception leaves open the possibility that the political could under certain conditions be made absolutely congruent with the ethical, optimism which I do not share” (ibid., 101 fn. 31). We need not comment here on her reading of Honig and Connolly, as it is not directly relevant (though this study does disagree with Mouffe on this point; see 6.7.2). What can be said, however, is that Mouffe rejects any approach to the political that even hints at the possibility of society reaching agreement on ethical or moral questions. She regards such positions as expressing an “optimistic view of human sociability” (ibid., 130) that “denies the negative side of exchange, its dissociating impulse” (ibid., 131), although she does not fully explain how morality itself engenders this optimism.

4.6 TO CONCLUDE: AGONISM AND ITS DISCONTENTS

The discussion in this chapter has focused on elucidating the foundational components of agonistic thought – its ontological and epistemological commitments and the assumptions that underpin them – while placing less emphasis on the practical enactment of agonism, that is, on how theorists conceptualise and implement agonistic democracy. In doing so, it has also

examined the intellectual premises, conceptual tensions, and normative commitments that inform agonistic democratic theories.

The central idea of agonism shared by all agonist theorists is that political contestation of existing structures, orders, ideas, relationships, and arrangements is not only the default state of affairs in the social world but also a desirable one. Contestation, or what might be termed a *healthy conflict*, should therefore not be contained, but actively maintained.

Like political realism, agonism emerged as a critique of liberal theorising, which seeks moral closure – an outcome considered by both realists and agonists to be not only impossible, due to the incommensurability of foundational ethical beliefs, but also undesirable, since moral ideals, once crystallised, can easily serve to conceal undesirable practices or power relations. Unlike realists, for whom stability is a central concern and the purpose of politics is to achieve pragmatic order through a *modus vivendi*, with conflict regarded primarily as a threat to stability, agonists view politics as a domain of positive disruption, in which conflict is empowering and central to democratic vitality.

Accordingly, realism can be characterised as conservative, with citizens treated as relatively passive subjects of political order, whereas agonism is emancipatory, granting citizens active agency and requiring their sustained participation.

Agonism is deeply influenced by continental philosophical thought, including Nietzsche, Schmitt, Wittgenstein, Arendt, Foucault, Deleuze, and Derrida. This intellectual lineage underpins its anti-foundationalist orientation, rejecting comprehensive doctrines, teleological assumptions about the good life, substantive accounts of human nature, and conceptions of legitimacy grounded in consensus (Wenman 2013, 2, 4). Instead, agonism values not only pluralism but also the process of conflict itself, reflecting its Greek origin (*agon*, which translates as struggle).

The purpose of engaging in this struggle is to generate new ideas, perspectives, and ways of thinking by disrupting the existing order. Thus, the aim of agonism is not to achieve stability through determinate resolutions or clearly defined ends, but rather the opposite – to actively contest prevailing structures and commitments. Yet, this is not a call for destructive or nihilistic disruption. Agonism is concerned with productive disruption – with questioning whether the stability we experience arises because everything genuinely fits together, which is unlikely given the persistent diversity of human societies and their ever-changing structural circumstances, or because such stability is superficial, maintained to obscure underlying tensions and ensure compliance. Stability and compliance, in this sense, reinforce one another – stability depends on

compliance, and compliance perpetuates stability. However, this mutual reinforcement may be detrimental to the vitality of the social body, which risks stagnation if it is not continually challenged to adapt, adjust, and engage in the very struggle that keeps it alive. Hence, agonism seeks to destabilise the sphere of political action in order to revitalise and transform it.

Bonnie Honig's conception of agonism positions itself against what can be described as moralist theorising, which understands the purpose of politics as securing a stable order through conceptual closure – whether in the form of a Kantian-Rawlsian consensus or Sandel's communitarian cohesion. In both cases, the desire for conceptual determinacy is, according to Honig, problematic, as it compels political subjects to suppress their creative and aesthetic drive towards positive contestation. For Honig, such contestation serves to open a sphere of reflexivity through critical dialogue – by engaging with perspectives different from our own, we come to understand the grounds of our commitments more fully and may even revise them. Moralist conceptual closure, by contrast, displaces the political process, transforming it into a sphere of truth-affirmation rather than reflexivity and adaptation.

However, Honig does not reject the importance of consolidation-seeking altogether. Instead, she views it as a necessary element of agonistic reasoning, since the pursuit of stability produces pressure points that subsequently become the sites to which disruption can respond. Nonetheless, Honig urges us to focus on developing agonistic rather than consolidatory thinking, as she considers the former to be more reflexive and self-correcting. The stabilising impulse, she suggests, arises naturally, whereas agonistic disruption requires deliberate cultivation.

Consequently, Honig foregrounds the existential and aesthetic dimensions of disruption, emphasising the creative potential of renewal. For her, resistance and nonconformity are not pursued merely to oppose others or to enhance one's social standing, but to create spaces of engagement with the 'other' that enable deeper self-understanding. Through this process, the search for common ground – stability itself – becomes not an aim of politics but a means of sustaining reflexive and dynamic democratic life.

William Connolly's central argument also stands against closure-seeking moral reasoning, which inevitably becomes doctrinal, generating unhealthy resentment that, in turn, fuels conflict and violence. To counter this tendency, Connolly advocates for critical responsiveness and agonistic respect as dispositions that enable political agents to recognise the contestability of their own commitments while affirming the legitimacy of others' struggles. However, unlike Honig, whose focus lies on rupture, Connolly's conception of agonism is

more concerned with the individual agent, emphasising responsiveness and coexistence amid difference. Accordingly, Connolly underscores the importance of acknowledging the ontological pluralism of political agents and, from this standpoint, highlights the individual responsibility for reflexive engagement within the social sphere.

Chantal Mouffe's agonism is the most explicitly political one, emphasising collective agency and the radical mobilisation of relationships that have been suppressed or ignored (e.g., those based on gender or ethnicity). This suppression stems from the same pursuit of closure critiqued by Honig and Connolly – the drive for consensus, whether (moral) substantive or procedural – which represses antagonism, which Mouffe regards as inherent in the social world. Yet for Mouffe, antagonism is not purely negative; it also contains a positive dimension, functioning as an affective drive that motivates mobilisation. Politics, therefore, seeks to transform the potential of antagonism into a productive force for civic agonism, first by recognising foundational differences and then by addressing the tensions created by suppression through the hegemonic construction of a shared symbolic space.

For Mouffe, substantive consensus is impossible and foundational moral commitments are irreconcilable; they cannot be suppressed but must instead be recognised and channelled towards a shared civic framework. This position presupposes that political agents embody the shared ethos of liberty and equality, which Mouffe regards as essential for agonism to operate – the subject must, in other words, possess an agonistic disposition, a demand which may be difficult to sustain within contexts shaped by entrenched asymmetries of power and the hegemonic forces that work to suppress such dispositions.

Mouffe's approach operates primarily at the level of institutional design and hegemony. In contrast to other agonistic theorists who emphasise disruption, she foregrounds stabilisation, prioritising the construction of common civic ground over the disruptive potential of agonism itself. This position seems to bring Mouffe closer than she might wish to the Kantian and Habermasian rationalists she critiques. The reason for this is her conviction that “every order is a hegemonic order” (2013, 14); consolidation, therefore, is not only inevitable but also constitutive of the political – it is merely a question of what kind of hegemony is established.

In sum, synthesising the insights of these three agonist thinkers – whose approaches are largely complementary, each emphasising different dimensions of agonistic thought – helps to clarify the ontological and epistemological foundations of agonism. Importantly, agonism fundamentally rejects both ontological and epistemological essentialism – that is, the notion that existence,

and our knowledge of it, are fixed and independent of observation. Instead, it embraces an ontological pluralism that rejects the existence of any universal foundation upon which perspectives – both moral and non-moral – can be reconciled. Conflict is thus understood as constitutive of the social world, and *agon* (struggle) is not external to politics but constitutes its very condition of possibility. From this perspective, the pursuit of comprehensive truths and consensus is detrimental to political life, as it produces a form of stability that amounts only to a superficial order maintained through the exercise of power. Likewise, rationality is conceived as a tool of power, operating through discursive practices, whereas political processes inherently involve creative (aesthetic) and affective dimensions that animate interactions among diverse identities, symbols, and cultural expressions – thereby generating the language of agonistic engagement. This language fosters reflexivity and recognises truth claims as interpretive rather than universal, so that understanding – first of oneself, and subsequently of others – emerges through meaningful engagement with difference.

Yet agonism is not without its weaknesses. Crucially, it assumes that political actors possess an inherently agonistic disposition – that is, a willingness to respect, rather than merely tolerate, the moral equality and liberty of others. Agonism presupposes that participants value engagement in constructive conflict as a means of fostering reflexivity and collective understanding, even above the pursuit of their own sense of justice, which is often rooted in deeply held beliefs rather than civic roles or interests. When actors are unwilling to acknowledge their own fallibility, conflict is unlikely to be productive or self-correcting. In such cases, efforts to cultivate civic agonism may be co-opted or ignored by those who prioritise affirming pre-existing moral or ideological commitments. As a result, agonism overestimates the normative capacities of political subjects, failing to account for those who lack the disposition necessary for the constructive interplay of difference.

Agonism nonetheless comes closest of the three approaches discussed here to escaping the priority of value over valuation because it refuses to settle the content of the good and treats political contestation as constitutive of politics rather than merely as a means of resolving problems. Yet the framework within which contestation takes place still rests on values settled in advance. For Mouffe, this entails the ethico-political principles of liberty and equality; for Connolly, the ethos of critical responsiveness; and, for Honig, the willingness of adversaries to treat contestation as sparring rather than as contest. I do not wish to argue that these commitments are mistaken (this is not relevant here); my point is merely that these accounts do not explicitly explain how the dispositions they require of

political actors are formed. Instead, they treat these dispositions as outcomes of the very practices that they are also supposed to make possible.

5. VIRTUE AND ETHICS

Having surveyed the conceptions of moral and ethical thought, together with the three strands of political theorising, each with its own way of accommodating social diversity – for some an inevitability, for others a welcome feature of social life – we are now closer to the proposal of this study, which develops an ethos of creative valuation based on virtue ethics and bridges it with normative theorising about politics.

Before setting it out in the next chapter, however, I want to discuss three existing accounts of the standing and function of virtues in ethical thought, from each of which I inherit something while rejecting its central proposal: Aristotle's teleological eudaimonism (5.1.1); Slote's sentimentalism (5.1.2); and Swanton's target-centred ethics (5.1.3). I then turn to Nietzsche (5.2), on whom this study draws heavily, though he is not traditionally counted a virtue ethicist. Finally, we take a brief look at how virtue ethics and political theory have related to one another thus far (5.3), before concluding with an overview of this chapter (5.4) that emphasises the immoralism of virtue.

5.1 ON VARIATIONS OF VIRTUE ETHICS

5.1.1 Aristotelian virtue ethics

Aristotle's account has shaped much of what followed in virtue ethics, and while this alone would make his work a good starting point, I will discuss Aristotle to clarify what my own account inherits from his and where it departs. Aristotle establishes the end first and understands the virtues as contributing to its attainment, whereas my account in chapter 6 rejects any telos and considers virtues as means of meaning-making. What follows, then, is neither a survey of Aristotle nor an adjudication between competing interpretations of his work; rather, I trace his teleological eudaimonism in the *Nicomachean Ethics*, with some reference to the *Politics*, in order to provide a background against which my own account can be situated.

Aristotle's worldview is grounded in the commitment that every thing has a function (*ergon*), and that things achieve their good by performing that function well. The *Nicomachean Ethics* opens by observing that every craft, inquiry, action, and deliberate choice “seems to seek some good” (NE I 1 1094a1-2). The connection between function and goodness is then drawn through the example of craft. Just as for a flute player “the good – the doing well – seems to lie in the function”, so too, Aristotle argues, this must be true of a human being, “if indeed there is some function that is his” (NE I 7 1097b25–27). This analogy tells us that the human good

does not arise accidentally or *ex nihilo*; it must be realised through practice. I would extend this thought by arguing that any serious practice is also a matter of experience, since experience is what makes a practice one's own, whether playing the flute or doing good. Activity by itself may be merely mechanical mimicry to which the agent attaches little or no significance, while experience, by contrast, is what gives practice its meaning.

I retain the Greek *eudaimonia* throughout, while Reeve, whose translation I use here, renders it as 'happiness' and defends that choice against 'flourishing' (2014, liv–lv). There is, however, no settled agreement on the most suitable translation. 'Happiness' seems to be used most widely in translations of the *Nicomachean Ethics* (Vittersø 2025), with 'flourishing' having become close to standard in neo-Aristotelian writing (e.g., Foot 2001), while 'living well' and 'well-being' also circulate, and Hursthouse treats *eudaimonia* as encompassing all three (1999, 9–10). Each of these translations, however, carries an account of what the end consists in, since 'flourishing' suggests a naturalist one, while 'happiness' suggests a psychological state. My objection, however, is precisely to the priority given to the end rather than to its content. Aristotle's own formulation – that to be *eudaimōn* is to live well and to do well (NE I 4 1095a19–20) – leaves the question open. Since I take this openness to be important, I retain the original Greek term.

What matters most for my purposes, however, is that *eudaimonia* is an activity (*energeia*) and not a state, so that it cannot simply be possessed but must be exercised: "the human good turns out to be activity of the soul in accord with virtue" (NE I 7 1098a16–17). Aristotle then asks whether the best good consists "in virtue's possession or in its use", and answers that one may possess the state while achieving nothing at all, "for example, if he is sleeping" (NE I 8 1098b31–1099a1).

What, then, are these virtues? In terms of their role in reasoning, virtues operate on two levels. First, Aristotle identifies the human function without any appeal to virtue, by asking what activity is peculiar to us and eliminating nutrition and perception, which we share with plants and animals. What remains is "activity of the soul in accord with reason" (NE I 7 1097b33–1098a8). Only then does virtue enter, as something that enables the activity to be performed well (NE I 7 1098a16–17). Virtue is thus defined by reference to the function, akin to the way the excellence of an eye is defined by seeing and that of a horse by running (NE II 6 1106a15–24). Second, in practical deliberation, virtue enables an agent to see the end correctly, since "virtue makes the target correct, and practical wisdom what furthers it" (NE VI 12 1144a7–9), and someone with a poor character will have a distorted view of what is worth pursuing.

In terms of kinds of virtues, Aristotle distinguishes two: virtue of thought "both comes about and grows mostly as a result of teaching ... which is why it requires

experience and time”, whereas virtue of character “results from habit (*ethos*)”, from which its name derives (NE II 1 1103a14–18). Of the intellectual virtues, he names five: craft knowledge (*technē*), scientific knowledge (*epistēmē*), practical wisdom (*phronēsis*), theoretical wisdom (*sophia*), and understanding (*nous*) (NE VI 3 1139b15–17). *Sophia* is the excellence of the contemplative activity that he will rank highest.

Of virtues of character, Aristotle says that we are “naturally receptive of them and are brought to completion through habit” (NE II 1 1103a23–25). This habit, in turn, “we acquire by first engaging in the activities” (NE II 1 1103a31–32), just as we become lyre players by playing the lyre, and “similarly, then, we become just people by doing just actions” (NE II 1 1103a26–b2). Habituation is therefore what makes virtue of character a stable state (*hexis*). Yet stability is not the same as crystallisation. What habituation establishes is a standing disposition, activated by choice and responsive to context, since virtue is “a deliberately choosing state, which is a medial condition in relation to us”, defined by reason and “by which a practically-wise person would define it” (NE II 6 1106b36–1107a2).

Aristotle argues against what contemporary political theory could term a *modus vivendi*, saying that the polis is not merely an object of a treaty of non-aggression or commerce; otherwise, “the Etruscans, the Carthaginians, and all those who have treaties with each other would be like citizens of one city” (Pol. III 9 1280a35–38). The treaty between Etruscans and Carthaginians governs how they treat one another, but says nothing about the kind of people they are. The law is a guarantor of just claims between people, which Aristotle allows as far as it goes but rejects as an account of the city, because such law is “not such as to make the citizens good and just” (Pol. III 9 1280b10). An association of that kind, he adds, differs from an alliance only in that its members live in the same place. What makes the difference is that “the city must be concerned about virtue” (Pol. III 9 1280b6). This is why the polis “comes to be for the sake of living, but it exists for the sake of living well” (Pol. I 2 1252b29–30).

How, then, can the polis be ‘concerned about virtue’? Aristotle tells us that knowing about virtue is not enough (NE X 9 1179b1–2), and that “ordinary people obey force rather than argument” (NE X 9 1180a4–5). Simply instructing citizens to be good would therefore be insufficient. What is required is for habituation to continue into adult life by means of law: “and so there will need to be laws concerning these matters as well and, in general, then, concerning all of life” (NE X 9 1180a2–4). Aristotle continues that “the law does have the power to compel, being reason that derives from a sort of practical wisdom and understanding”, and, unlike a person who opposes our impulses, “the law causes no offense in prescribing what is decent” (NE X 9 1180a20–24).

From this reading we see that Aristotle does not extend an ethics to politics, as for him they were never apart. Virtues do not emerge from nowhere, nor are they taught as a doctrine is taught. They arise from acting among others, in settings arranged so that the right actions become habitual, which is why “it is from acting as we do in our transactions with other human beings that some of us become just and others unjust” (NE II 1 1103b14–15). We cannot, in short, become virtuous alone. What the public sphere supplies is the framework for practices in which virtuous judgements can be formed, and this is the sense in which “legislators make citizens good by habituating them” (NE II 1 1103b3–6).

I will now set out what I set aside and what I can retain from Aristotle in developing my proposal in chapter 6.

I set aside the teleological order of reasoning that places the end prior to the virtues, whereas my own Nietzsche-inspired account does not treat virtues as means towards an end specified in advance, but as dispositions through which values acquire their meaning, relevance, and significance. It may seem that Aristotle stands closer to Nietzsche than this contrast suggests, since Aristotle conceptualises eudaimonia as an activity rather than a fixed state, but this resemblance is misleading. As I discuss in 5.2.2, Nietzsche resists fixing any state of well-being as a form of “being” – and places becoming in its stead ([1882] 1974, 266). Aristotle also refuses to treat living well as a state, but for a different reason, arguing that it is possible to possess virtue while inactive or even asleep (NE I 5 1095b32–1096a2). For Aristotle, the relevant activities are complete in themselves. As he puts it, “seeing seems at any time to be complete, since it does not lack anything” whose later coming to be would complete its form, whereas “every process (for example, building) is in time and relates to some end” (NE X 4 1174a14–20). The openness of eudaimonia, then, does not mean that one develops towards some later or higher state. It means that the activity is repeatedly enacted and, through repetition, becomes a more or less stable habit. For Nietzsche, by contrast, incompleteness is constitutive, because the higher state never experiences itself as having arrived, and once it does, its virtues crystallise and cease to be virtues ([1882] 1974, 266). Thus, both deny that living well is something one can possess, but for opposite reasons – Aristotle denies it because he views activity as complete in itself, whereas Nietzsche sees it as an ongoing process of becoming that can never be complete. Nietzsche, as we shall see, does not begin with any fixed determination of what one is or should become. For this reason, *becoming what one is* cannot mean actualising something that already exists within us, as the *what* is configured and created rather than discovered (Dennis 2019, 62–63).

There are, nevertheless, three points at which Aristotle’s account is compatible with my proposal. First, like the broader virtue-ethical revival, my account treats

ethics and politics as belonging to the same field of inquiry, which is precisely what the moralists, realists, and agonists of chapters 2–4, in different ways, seek to pull apart. Second, Aristotle places practical judgement in the person rather than in a universal principle, arguing that what is appropriate is relative to the agent and to what the action concerns (NE IV 2 1122a24–25). This is important because, as Aristotle observes, noble and just things “admit of so much difference and variability” that they can appear to exist by convention rather than by nature (NE I 3 1094b14–17). Thus, Aristotle does not make the person a creator of value in my sense, but his emphasis on situated judgement and the variability of virtuous action opens up a space for an epistemic account of value – when Aristotle’s wise person determines what virtue requires in a particular situation, they do not determine what counts as good in the first place. And such a conception of the person as the locus of practical judgement is what my framework can accommodate. And third, Aristotle defines virtue and practical wisdom (*phronēsis*) in relation to one another, arguing that it is “neither possible to be fully good without practical wisdom nor practically wise without virtue of character” (NE VI 13 1144b30–32). This aligns with my own account (chapter 6), which treats virtues, dispositions, and social values as interconnected; for example, the virtue of honesty cultivates a disposition towards truthfulness, which in turn gives meaning to the social value of trust. In moral terms, honesty is understood as a principle; in ethical terms as a virtue.

In this respect, my view also aligns with Aristotle’s idea that ethical values (i.e., what we find valuable for our soul, to use Aristotle’s terms) do not exist prior to social relations: “it is from acting as we do in our transactions with other human beings that some of us become just and others unjust” (NE II 1 1103b14–15). I therefore distinguish between virtues and values, where honesty is a virtue rather than a value because value does not exist independently of valuation. It gets substance (meaning for the valuer) through activity performed by the valuer in the particular situation. In this sense, value is not something that exists prior to the exercise of virtue but is constituted through it, a point that Aristotle’s account also helps to illuminate.

5.1.2 Agent-based sentimentalism of Michael Slote

Michael Slote is the leading proponent of agent-based virtue ethics, which emphasises the agent’s motives. “[A]gent-based virtue ethics ... understands rightness in terms of good motivations and wrongness in terms of the having of bad (or insufficiently good) motives” (2001, 14). Similarly, Linda Zagzebski defines the wrongness of an act from the agent’s perspective, in that they “would feel guilty if [they] did [wrong]” (2005, 160, emphasis added). Thus, in an agent-based account, it is the agent who determines what is virtuous, guided by their own

interpretation – rather than rationalisation – of the good. Virtue is therefore understood as the motivation characteristic of an exemplary agent (Hursthouse and Pettigrove 2023), namely, a sincere intent to do good. While the action matters, its outcome is secondary, as its moral significance derives from the virtue motivating it, not from its consequences.

This motivational focus becomes problematic when virtue is viewed holistically. If virtue is equated solely with motivation, one risks implying that any action done with good intent is virtuous. Yet it is perfectly possible to act wrongly for the right reasons. Benevolent intent may be a moral motive even when it leads to harmful outcomes, but the reverse is counterintuitive – for example, treating the accidental saving of a life as a virtuous act. Although saving a life is the right action, Van Zyl (2009, 2011) argues that it is not necessarily morally praiseworthy, since praiseworthiness presumes that the agent intended the good.

This raises a further question: by what mechanism are traits judged virtuous? How does an agent determine what is good? Slote (2001, 2007, 2010, 2013, 2023) develops a sentimentalist approach inspired by early sentimentalists such as Hume, while rejecting his epistemological scepticism and engaging instead with care ethics. However, as will be shown, Slote's (2001, 93–102) approach – although virtue-ethical and rejecting pure rationalism, in line with the stance of this study – ultimately fails to reconcile personal sentimentalist reasoning with public reasoning, particularly regarding justice. This failure arises because it seeks, alongside both care ethicists and moralists, to articulate a universally acceptable conception of justice – an endeavour which, as noted in section 1.4.1, is bound to fail.

Slote's sentimentalism locates moral motivation primarily in emotion rather than abstract rationality. He identifies empathy as the central moral capacity (2010, 13–25), treating it as fundamental to moral epistemology (2013, 13–14). Importantly, Slote does not regard empathy as an emotion in the narrow sense – an involuntary affective response – but as a complex blend of psychological, cognitive, and emotional capacities enabling understanding and responsiveness to others.

He argues that “objectivity and open-mindedness with respect to ideas, facts, and/or arguments are basically and solely a matter of being empathic in certain (as yet here unspecified) ways” (2013, 40). Knowledge, therefore, concerns the reasons underlying one's ethical commitments, and genuine understanding of another's moral stance requires empathy. Factual knowledge – “unspecified” prior to inquiry – demands open-mindedness and a willingness to move beyond one's existing beliefs. Empathy, Slote claims, uniquely fosters this openness. Thus, epistemic rationality, understood as the genuine desire to acquire

knowledge, requires empathy. Empathy is therefore a basic component of thought and cognition, not merely an affective phenomenon to be restrained or suppressed.

Empathy also prevents one's attitudes towards others' beliefs from sliding into hostility. Tolerance, then, is not merely pragmatic civility but a coherent disposition towards difference sustained by empathy:

“Since anger and hatred toward those one disagrees with intuitively seem incompatible with being objective about their opinions, it would seem that a degree of empathy incompatible with great anger and hatred is a necessary condition of being objective or open-minded in regard to the viewpoints or opinions of others” (Slote 2013, 40).

Slote denies that incorporating an emotive component into epistemic rationality makes it irrational. Instead, he argues that rationalism – such as Kantian rationality – permits the possibility of rationally acting immorally (2010, 97–106). Truly moral action requires more than rational deliberation; it requires sentiment. As he writes:

“Total immorality entails heartlessness, not irrationality ... the sentimentalist regards nonempathic psychopaths as incapable of understanding moral claims or making their own moral judgments ... such claims or judgments ... cannot give the psychopath reason (or motivation) to act.” (2010, 51 fn. 12)

Immorality thus stems from lack of empathy, not from irrationality or ignorance. Moral claims can be rationally justified, but this justification provides no genuine motivation for the non-empathic agent, who cannot grasp the sentimental dimension of morality.

Making sentiment primary and rationality secondary implies that moral reasoning still functions where rationality alone cannot supply sufficient knowledge for decision-making. This is not to say that a mere ‘gut feeling’ can resolve ethical dilemmas. Rather, while rationality stripped of sentiment may lead to immoral choices, sentiment without rationality is comparatively better placed to facilitate morally sound action.

Slote also urges rejection of the ideal of perfection – understood as the full realisation of human capacities – and the rationalist belief that life should follow a fixed path towards completion. The rationalist desire for control, he argues, reflects a failure to recognise “receptivity to life” (2013), understood as an open, flexible attitude that embraces what comes and adapts to circumstances.

“Enlightenment-type ethical rationalism and epistemological rationalism ... express an inordinate desire for control ... unwilling or unable to acknowledge ... the value of receptivity and of emotions that involve it.” (ibid., 202)

This receptivity contrasts with the idea of a “life plan” (Slote 2013, 169–175), which Slote attributes to Rawls, drawing on Aristotle. A life plan organises one’s existence around coherent long-term goals expressing one’s character and values. Moral development, in this view, involves shaping desires and choices in accordance with such a plan: “What one does according to the plan eventually makes a difference to the kind of person one is” (ibid., 172).

Slote associates this telic rationality with liberal thought, which he sees as oriented towards controlling one’s choices in line with a preconfigured path. This does not reject particular aims but seeks to regulate life through a fixed purpose. Such constant rational evaluation risks neglecting aspects of life that resist this mode of assessment and forecloses possibilities that a receptive disposition would otherwise allow.

He argues that the “modern-day liberal idea that we ought to subject all our feelings, assumptions, and relationships to critical rational scrutiny” (Slote 2013, 18) means that individuals may misinterpret “things that life *has brought them*” (ibid., 177, original emphasis). Elements of life that do not fit within a life plan should not be evaluated in the same way as those that do. Constraining epistemic openness to the unexpected risks missing opportunities to enrich one’s life.

To summarise, Slote advances a sentimentalist account of justice as an alternative to Rawlsian liberal moralism. While he rejects strict rationalism and makes empathy the basis of both moral motivation and moral knowledge, he retains a universalised conception of justice, which creates a tension within his account, since a universal conception of justice requires abstracting away from the particularity of empathy. This study therefore shares Slote’s emphasis on the sense of justice, while setting aside its universalist ambition.

5.1.3 Pluralist and target-centred virtue ethics of Christine Swanton

Having discussed two established perspectives on virtue ethics – the eudaimonist approach, which adopts an ends-oriented view of virtue, and the agent-centred perspective, which defines virtue primarily in terms of motivation – we now turn to a more recent account proposed by Christine Swanton (2003, 2021).

Swanton presents her view in opposition to moral or ethical theories that offer specific guidelines for action. As she puts it, “I wholeheartedly reject the idea that

it is the purpose of an ethical theory to ‘tell us what to do’ in our quest for answers to specific moral questions” (2021, 11). Instead, her target-centred ethics provides a framework – or map – for understanding the ethical domain, giving it an inherently critical function. A critical framework that accommodates ethical complexity, she argues, enables us to interrogate our own ethical understanding, which is always shaped by theoretical and cultural assumptions. By contrast, simplistic theories such as monistic consequentialism, which judge everything solely by outcomes, distort ethical reasoning and risk overlooking essential moral dimensions, including motives and emotions (ibid., 12).

Swanton conceives virtue as a disposition that enables individuals to respond to ethical demands while taking into account multiple, often competing, considerations. She thus rejects the ideal of moral perfection. She defines virtue as “a good quality of character, more specifically a disposition to respond to, or acknowledge, items within its field or fields in an excellent or *sufficiently good manner*” (2003, 19, emphasis added). This definition neither treats virtuous acts as those that achieve the optimal response nor seeks to establish a fixed or final standard of virtue. We cannot, therefore, claim that honesty or courage is inherently virtuous. Rather, Swanton maintains that “the standards for virtue are relative to context” (ibid., 24). Because “the chances of us developing to perfection the psychological basis for leading the best life are vanishingly small” (ibid., 14), virtue must be understood as context-dependent and open-ended. For Swanton, the moral landscape is too complex – with competing values and overlapping obligations – for any ideal of perfected virtue; some degree of deficiency is inevitable, yet such lack is not equivalent to loss (ibid., 201). Deficiency signals room for growth, whereas loss implies something irretrievable. Virtuous action, therefore, is not about reaching final moral resolution but about remaining open to continual refinement in response to changing circumstances.

Thus, Swanton’s account is target-centred because she sees virtues as directed towards context-specific demands. The “target” is set by the needs of others, the situational context, and the moral imperatives at stake (2003, 232–238). A virtue’s function is to elicit an appropriate response; virtues are judged by how well they meet their moral target, not by conformity to fixed rules or stable traits. Courage, for example, is virtuous not because it follows a general rule, but because it meets the demands of risky situations without collapsing into recklessness.

Moreover, a virtue’s target is not straightforward. A virtue does not aim at a single, uniform response. Instead, hitting its target may require multiple modes of response – for example, justice may call for both respect and restraint. Targets may be internal as well as external, requiring the cultivation or overcoming of

inner states such as fear or bias, not merely outward action. There may also be several possible targets in a given situation, which have to be weighed against one another. Context also affects what it means to hit the target, shaping what is considered relevant and what relationships are involved. Finally, virtue may call for inaction rather than action, especially when acting would do harm.

Virtuous responsiveness therefore does not involve applying a universal rule, but rather attending to the specific features of each moral situation. Swanton notes that even the apparently simple metaphor of “hitting the target” is complex. She identifies five factors, namely that a virtue’s target may require multiple modes of response; the target may be internal; there may be more than one target; context shapes the target; and the target may involve avoidance (2003, 233–238).

Consider workplace harassment. The situation involves inappropriate conduct; the target is the reporting of that conduct; and the relevant virtue is courage, as non-anonymous reporting carries a risk of retaliation. The response may involve at least two moral modes, namely respect for human dignity, and the promotion of well-being. The target may also be dual, either external (the act of reporting) or internal (overcoming fear of victimisation). Context further shapes the appropriate response – depending on the gravity of the misconduct, reporting might remain internal to a small group or be made public. Finally, hitting the target of courage may sometimes require avoidance – for instance, refraining from public disclosure not out of cowardice but from legitimate concern for others’ welfare.

Taken together, from Swanton’s perspective, virtuous action is rarely grounded in a single moral reason. Rather, it emerges from a constellation of interrelated reasons that influence and constrain one another. Virtue, on this view, is dynamic, context-sensitive, and open to continual revision. Virtue therefore does not determine the reasons, situation, target, or appropriate action. Rather, the agent’s attention is directed towards the situation itself and what it calls for. The focus is not on identifying or applying a virtue such as courage, but on recognising the reasons present in the situation that justify and guide an appropriate response.

When characterising virtue, Swanton treats virtues as “thick ethical concepts” – such as kind, generous, patient, or courageous – rather than “thin concepts”, which express general principles like obligation (2021, 1). Thick virtues (and vices) are evaluative, while thin concepts are abstract and action-guiding. Understanding virtue as a thick evaluative concept clarifies how virtues give meaning to the value of “hitting the target”, whereas thin principles keep action on the right track.

Thus, for Swanton, virtues are primarily creative, not merely practical, unlike in the Aristotelian account. This reflects her frequent engagement with Nietzsche, particularly his view that ethical reasoning is fundamentally a form of creation rather than practice (2021, 2). Since this Nietzschean conception of virtue is central to the broader aim of this study – to develop a virtue-ethical framework for political theorising – it is now appropriate to examine it in detail.

5.2 NIETZSCHE AND VIRTUE ETHICS

In section 1.3, we examined Nietzsche's critique of modern rationalist morality – a view echoed by political moralists, realists and agonists, all of whom consider morality to seek epistemic closure through conformism, universalism, and abstract generalisation. We can now turn to Nietzsche's positive ethical outlook, which many (Hunt 1993, [1991] 2005; Brobjer 2003; Daigle 2006; Swanton 2005, 2015; Dennis 2019) characterise as a form of virtue-based ethics – a categorisation this study also adopts.

Not all scholars, however, accept Nietzsche as a virtue ethicist. At the same time, it may be more accurate to view this as an ongoing debate over the extent to which he can be considered a normative virtue theorist, particularly of an Aristotelian kind (Magnus 1980; Leiter 2024). Although Nietzsche shares certain ideas with Aristotle (Hunt [1991] 2005; Daigle 2006), he rejects both *telos* as the ultimate aim of the virtues and the notion of universal virtues (Swanton 2015, 157). Yet, like Aristotle, he emphasises self-realisation as human excellence and a thin conception of human nature (*ibid.*, 158).

Overall, the vast body of Nietzsche scholarship shows that interpreting his work is far from straightforward. Developing a new interpretation within this study would be unfeasible, as it would require a separate project. It is therefore prudent to draw on a limited number of sources and to ground the analysis in carefully selected scholarship on Nietzsche. Most importantly, this study relies on interpretations of Nietzsche's ethics as qualitative rather than consequentialist, as advanced by Donovan Miyasaki (2022a) and Christine Swanton (2005), supplemented by insights from other scholars.

5.2.1 The foundations of Nietzsche's ethics

We have already seen how Nietzsche rejects morality by tracing its historical origins (1.3). We now turn to the foundations of his immoralism, which intersect with his account of the ethics of being human. To do so, we move from an examination of Nietzsche's (non-)nihilism to a consideration of the concept of *amor fati*, or the love of fate.

Existential (non-)nihilism

Nietzsche is often portrayed as an existential nihilist (Hatab 1987, 91). However, although his writings contain many references to nihilism (ibid.), these are largely critical (Brobjer 2003, 76–77 fn. 11). This raises an obvious question: are Nietzsche's own views nihilistic, and if so, in what sense?

To address this, it is helpful to begin with nihilism itself. In its strictest form, nihilism denies that human life has inherent meaning, claiming that existence has no purpose or significance, and thus no truth about it can be affirmed (Metz 2023). Yet nihilism cannot be an ontological doctrine, since purpose, relevance, and truth are not ontological categories. Relevance and truth are clearly epistemic notions, and so too, by implication, is the functional concept of purpose. Thus, nihilism is an epistemic position, holding that we cannot know things as they are in themselves, and thus cannot ascribe objective evaluative standards to them. For the nihilist, meaning and objectivity in relation to human life are mutually exclusive. Nevertheless, even for the nihilist, life persists, and its apparent meaninglessness can be understood as a form of freedom – one may succumb to despair or seek to fill the void through one's own creations. And this is precisely what lies at the heart of Nietzsche's ethical project. Let us now explore this in greater detail.

Hatab (1987) and Brobjer (2003) both note that Nietzsche does not reject all meaning. Rather, he rejects determinateness, including moralist conceptions of 'truth'. As he writes: "our opinions about 'good' and 'noble' and 'great' can never be proved true by our actions because every action is unknowable" (Nietzsche [1882] 1974, 265). This challenges consequentialism – since we cannot know the future – and duty-based ethics, which depend on intentions that are ultimately unknowable, our own included. Nietzsche thus dismisses the supposed authority of convictions and rational argument. What we take to be rational, truth-apt conclusions are, in the end, an illusion of our own making (Brobjer 2003, 64).

Nietzsche is therefore not a nihilist about value or meaning, but about determination, as he rejects predetermined categories. Far from denying value, he treats it as fundamental to human existence. This is clear in *Zarathustra*:

"Humans first placed values into things, in order to preserve themselves – they first created meaning for things, a human meaning! That is why they call themselves 'human,' that is: the esteemer. Esteeming is creating: hear me, you creators! Esteeming itself is the treasure and jewel of all esteemed things" (Nietzsche [1883] 2006, On a Thousand and One Goals).

Here, Nietzsche emphasises that the creation of value is what makes us human, for it is through this process that meaning comes into being. Human beings are, therefore, creators of meaning. Since the world lacks inherent significance, nothing external – moral or otherwise – can confer meaning on our existence. Meaning arises only through our own acts of valuation, which constitute the core of human experience.

Love of fate and the affirmation of suffering through resistance

Nietzsche argues that the moralistic (i.e., slavish) mode of thinking, with its reliance on fixed categories, produces a dualistic way of thinking in which moral judgements require one to decide whether an outcome is good or evil, thereby presupposing stark oppositions. He maintains, however, that good and evil are both matters of degree and mutually necessary (Brobjer 2003, 65). Suffering and happiness create the friction essential to development – personal, cultural, and social – because struggle is the site in which meaning itself is created. Yet this link between struggle and creation is not merely instrumental, as it can be for the Stoics. For Nietzsche, the affirmation of suffering through struggle is valuable in its own right (Miyasaki 2022a, 76). Whereas a Stoic might justify suffering as a means to a desirable end, such justification still treats meaning as something constructed to make suffering bearable. The lover of fate, by contrast, finds struggle intrinsically valuable and therefore lovable. Love, in this sense, is not a means to an end but an affirmation without the need for external validation or reward. Loving one's fate is not a matter of justifying or explaining it, but of actively embracing life as opportunity. Nietzsche's *amor fati* should thus not be mistaken for passive submission or reluctant acceptance.

This is clear in *The Gay Science*, where he first introduces *amor fati*:

“I want to learn more and more to see as beautiful what is necessary in things; then I shall be one of those who make things beautiful. *Amor fati*: let that be my love henceforth! I do not want to wage war against what is ugly. I do not want to accuse; I do not even want to accuse those who accuse. Looking away shall be my only negation. And all in all and on the whole: some day I wish to be only a Yes-sayer!” (Nietzsche [1882] 1974, 276)

Here, loving one's fate means embracing existence as it is. It does not involve striving for or against a predetermined destiny, but affirming fate as an essential aspect of life. The “looking away” signals the rejection of resentment and the affirmation of resistance – the willingness to say “Yes!” to fate, welcoming both joy and suffering, the latter as the precondition for the former. Suffering is necessary for a (playful) engagement with resistance, which in turn brings life its

joy. Drawing on Miyasaki's compelling account of amor fati as the non-instrumental affirmation of suffering (2022a, 100–115), we can further examine how Nietzsche links resistance, suffering, and the creation of meaning.

Underlying amor fati is Nietzsche's rejection of free will (Brobjer 2003, 65). This also leads him to deny that good judgement can arise from rational reasoning alone, since this would presuppose complete freedom of choice. The tension between moral obligation and the assumed freedom and rationality of the moral agent underlies Nietzsche's critique of modern morality (Miyasaki 2022a, 54–55). Traditional morality holds that norms, rules, and principles defining good and evil can guide individuals towards better actions. For Nietzsche, however, the rejection of free will and the limits of rationality mean that universal principles, deliberate choices, and rational reflection cannot by themselves transform one's character.

Thus, morality, or more precisely moralism – the claim that morality is objectively or universally justified and needs no further grounding – fails to bring about genuine transformation (Miyasaki 2022a, 55, 57, 59). Moreover, as discussed in section 1.3 on the genealogy of slave and master moralities, modern morality emerges from socially constructed inequalities – from political order itself. The belief that morality transforms people is therefore an illusion sustained by the political structures that promote it, reinforcing distinctions between good and bad actions and enabling power to operate through the internalisation of moral values.

Feeling of power and happiness through resistance

The rejection of free will and its rational exercise leads naturally to a concept of Nietzsche's closely linked to amor fati – the will to power (*der Wille zur Macht*), which can be understood as the driving force of creative living and self-overcoming. Miyasaki (2022a) distinguishes two interpretations of the will to power, namely quantitative and qualitative. Quantitatively, it is “the universal aim shared by every drive...the most basic human motivation...to increase power in relation to one's environment – its aim, in other words, is domination or conquest” (2022a, 77). This refers to measurable expansion of power over others. By contrast, the qualitative conception concerns the feeling of power, which Miyasaki identifies as Nietzsche's central concern rather than the accumulation of actual power (2022a, 77–88), aligning with Swanton (2005, 181–182), who likewise emphasises that Nietzsche rejects the maximisation of power.

This interpretation seems coherent if we accept Nietzsche's denial of free will – since no rational method exists to assert one's will, the will to power cannot aim at measurable acquisition. Such a desire would be inherently insatiable, as no

finite amount of power could ever suffice. Thus, amor fati entails accepting that there is neither endpoint nor measurable outcome capable of providing ultimate satisfaction. Consider the following passage from *The Antichrist*:

“What is good? – Whatever augments the feeling of power, the will to power, power itself, in man. ...What is happiness? – The feeling that power increases – that resistance is overcome.” (Nietzsche [1888] 1923, 43)

Here, the source of the feeling of power is the act of overcoming resistance. Suffering, therefore, is integral to loving one’s fate, as it emerges from experiencing obstacles, not merely recognising them rationally. Some resistance can be overcome, yet there will always be infinitely more, reflecting the impossibility of free will within Nietzschean thought. Consequently, the feeling of power drives the will to power, regardless of whether power actually increases or obstacles are fully overcome (Miyasaki 2022a, 86). As there is no objective free will, there is also no objective power (*ibid.*, 87). Power has no intrinsic meaning; it is the subjective experience – the *feeling* of power – that matters.

The Nietzschean ‘game of life’ is therefore sustained by confronting obstacles, since the effort to overcome resistance produces a feeling of power. These obstacles are not existentially necessary but are pursued for their own sake, as an expression of the will to power. As the passage quoted above tells us, happiness consists in the act of overcoming itself rather than in any particular condition it seeks to achieve. This feeling of power is intrinsically valuable, in the sense that it is non-instrumental, non-teleological, and non-quantifiable. Gaining power effortlessly would not generate joy, as it would not be the result of one’s own creativity and effort, rendering it fragile and illusory (Miyasaki 2022a, 78).

In contrast, pursuing quantitative power – such as through perfection or domination – ultimately undermines life. Eliminating obstacles removes the conditions for creative confrontation, resulting in apathy and emptiness. Ultimate power, achieved without resistance, becomes meaningless, dissolving the will to power. Nietzsche’s concept of the Last Man (*der Letzte Mensch*) in *Zarathustra* ([1883] 2006, Prologue §5) captures this well: frictionless happiness replaces striving, while power that is never tested by obstacles gives way to mediocrity (Kaufmann 1974, 162). Huxley’s ([1931] 2006) dystopian *Brave New World*, in which individuals are conditioned to feel contentment with their predetermined place in society, offers a similar parallel.

For Nietzsche, resistance is something to be preserved rather than eliminated, as it provides the joyful experience of playing the ‘game of life’ through the continual attempt to overcome obstacles. Were we to conceive power purely in

quantitative terms, the engagement with resistance would instead be motivated by the desire to eliminate obstacles (and suffering) altogether. If power is sought as a matter of accumulation, its ultimate form would be absolute power, which would in turn enable one's dogma to become universal, thereby creating the conditions required for sustaining that absolute power (Miyasaki 2022a, 77–78). Yet, because absolute power is impossible, a quantitative conception of power inevitably generates resentment, since one can never be satisfied if power is pursued solely in accumulative terms. This is precisely the problem with the slave (Judeo-Christian) morality discussed in 1.3, which inverts the evaluative table of 'good' and 'bad' – shifting it from strength to weakness, from domination to submissiveness, and so forth – because it seeks to maximise power while remaining unable to attain it within the framework of master morality. Consequently, slave morality reconceptualises the substantive content of power but not its quantitative conception.

In contrast, Nietzsche's feeling of power – or the creative drive of power – can flourish through *amor fati*, which allows one to conceive of power in terms of "contest, not conquest", as Miyasaki (2022a, 77) observes. To elaborate, loving one's fate presupposes acknowledging that we possess limits and encounter obstacles, both internal and external, without necessarily knowing what they are or how they will manifest. This absence of complete knowledge compels us to test and confront these limits. Such unknowability, and the corresponding persistence of indeterminacy within the love of fate, is precisely what enables the dynamic impetus of the will to power.

5.2.2 Nietzsche's ethics of virtue

Thus far, we have primarily examined Nietzsche's critique of morality as embodied in the Last Man of *Zarathustra*, in particular the moralistic, obligation-based, and solution-oriented modes of value-driven thought. Only briefly have we addressed Nietzsche's response to this condition. We now turn to Nietzsche's virtue ethics and to the figure of the Overman (*der Übermensch*), which stands as the positive antithesis of the Last Man and embodies *amor fati*, the will to power, and the affirmation of resistance.

As discussed earlier, Nietzsche's ethics – his account of what and why things are valuable, good, or bad in human life – differs markedly from ancient Greek virtue ethics. He rejects both the Aristotelian teleological framework and agent-centred approaches that privilege the intentions of the moral actor while treating action as the primary focus of ethical reflection (Leiter 1997).

Nietzsche may nonetheless be considered a virtue ethicist because of his emphasis on the human being as such – on character, dispositions, and the art of being

human. It is an art because, for Nietzsche, human beings are creators – and not only creators of themselves through self-mastery, but creators of value as such. Recall the passage from Zarathustra discussed in 5.2.1, which illustrates how Nietzsche locates valuation at the core of human existence: human beings “first placed values into things” ([1883] 2006, On a Thousand and One Goals), thereby creating meaning rather than discovering it. To be human is, therefore, to be an “esteemer” (ibid.) – a creator of value and, consequently, of meaning. For Nietzsche, the world lacks inherent or objective significance; meaning arises only through acts of valuation. This creative activity thereby renders human beings fundamentally aesthetic beings, in that value is affirmed not as a means to an end but for its own sake, as something intrinsically worth creating.

This also indicates that Nietzschean virtues do not constitute a list of traits associated with living a good life (such as honesty or courage), but concern the significance of traits as conditions that enable the creation of a life valuable to the creator. A trait that is virtuous in one context or for a particular individual may be a vice in another: “For Nietzsche, just as a universalist account of what is healthy for all individuals is unsustainable, so too is a universalist account of which character traits are virtues” (Dennis 2019, 58).

Whereas modern morality evaluates human beings in terms of their actions, Nietzsche values the human capacity for creativity and places the very being of the human person at the centre of philosophical reflection. He reframes the questions traditionally posed by philosophers – such as “What is truth?” or “What ought one to do?” – by recasting them in explicitly person-centred terms: “Who is to be created?” or “Who is most valuable?” (Brobjer 2003, 73). Questions concerning truth and appropriate action are not only oriented towards problem-solving and practical decision-making; they are also abstract in that they presuppose that truth and right action can be identified independently of the particular individuals who embody them. Nietzsche, by contrast, reconfigures these inquiries so that the person becomes their central focus.

Nietzsche rejects rationalistic accounts that treat reason as the primary faculty of creativity (Daigle 2006, 7). For him, rationality is the means that makes possible the justification of obligation and the guidance of action in the first place, including the rationalisation of intuitions undertaken by moral philosophers such as Henry Sidgwick and W. D. Ross. For Nietzsche, however, thinking about action prior to the actor is fundamentally misguided, because, for him, rationality constitutes only a minor aspect of our nature. Instead, Nietzsche describes the human being as a “social structure composed of many souls... social structures of drives and affects” ([1886] 1998, §19, §12), further noting that this structure resembles a labyrinth rather than a neatly organised and coherent

whole ([1881] 1997, §169). This labyrinth represents the complexity and disorientation of our inner selves. Any formal, general principle intended to guide our actions would first have to identify specific points within this labyrinth that could provide reasons and motivation for us to act. Yet, much of moral philosophy presupposes that an action can be moral only if it carries its own internal reasons – if moral action is internalised. Within a labyrinth, however, establishing such internal grounding through abstract universal principles becomes extremely difficult, if not impossible, because the reasons that move a particular agent – if Nietzsche is correct – are not solely rational. This reinforces the idea that morally right choices – those genuinely felt as right by the moral agent – are ultimately unpredictable.

The creative Overman

What kind of being, then, does Nietzsche have in mind when he speaks of the human? Who is Nietzsche's creator of value?

In *Zarathustra* ([1883] 2006), Nietzsche introduces the Overman (*der Übermensch*) as the positive antithesis of the Last Man discussed earlier. The Overman is a being that is more than human, but not more of a human. Nietzsche states that a "human being is something that must be overcome" ([1883] 2006, Prologue §3), signalling that the Overman should not be confused with the concept of the superhuman, who would embody a quantitatively maximised state of desirable human characteristics. The term 'over' signifies the overcoming of the confusion inherent in the labyrinth of the human soul ([1886] 1998, §19). Whereas the superhuman seeks a fixed, ultimate state of being, the Overman exists in constant flux, continually navigating the internal labyrinth, which is itself ever-changing due to the shifting drives and affects of human nature.

The Overman masters their internal labyrinth and does not seek to impose their choices on others. Rather than formulating principles that could be universalised – as deontological and utilitarian moralities require – the Overman "will be true to him- or herself" (Daigle 2006, 9), an ideal encapsulated in the notion of *becoming what one is* (Dennis 2019, 61–64)¹⁴. The Overman embraces *amor fati* and wills to power, creating themselves as "new, unique, incomparable, who give themselves laws, who create themselves" (Nietzsche [1882] 1974, 266). Since, as a default condition, "everybody is farthest away from himself" (*ibid.*, 263), the process of becoming begins with recognising that any self-conception not created by oneself is an illusion. Nietzsche warns against being enslaved by such illusions – an

¹⁴ Originally from *The Gay Science*, translated also as "become those we are" (Nietzsche [1882] 1974, 266).

enslavement that arises when we adopt, follow or imitate others without asking where our conceptions of the right come from (ibid., 263–264).

Dennis (2019, 62–63) identifies three themes in Nietzsche's account of 'becoming what one is'. First, uniqueness results from the particular configuration of an individual's drives as they interact with their environment. Second, one must attend to one's conscience by interrogating the conscience behind the conscience – that is, by critically examining one's reasons (ibid., 62). Third, one requires the courage to embrace unconditional honesty, confronting realities that lie beyond the habitual and conventional. Dennis further argues that such *becoming what one is* constitutes Nietzsche's conception of human flourishing (ibid., 64).

Two points are especially significant. First, becoming – rather than being – captures Nietzsche's view of the self as a condition of continual flux and transformation (Nehamas 2001, 260–262). Second, this process requires distinguishing oneself from others, so instead of seeking external validation or reproducing established models of the good, the Overman generates values from within. Creative activity thus cannot amount to imitation – whether of others or of one's earlier self – because each individual's configuration of drives and circumstances is irreducibly unique.

In short, the Overman is not concerned with the social dimension of human interaction but with the task of self-cultivation. This self-cultivation is not a matter of developing or enhancing virtues so that they become habitual and useful for action; rather, for Nietzsche, virtues are both constitutive of 'becoming what one is' and expressive of one's unique endowment (Dennis 2019, 71). Virtues are neither merely instrumental to flourishing nor universal across individuals, but integral to the Overman's transformative process. This process involves a movement from a lower state characterised by submission, mediocrity, powerlessness, and resentment, towards a higher state of self-mastery – a "continuous matter of overcoming obstacles", in which "goals may themselves change as one becomes stronger and faces new obstacles and circumstance" (Swanton 2005, 189). The higher state never experiences itself as having arrived; the lower state feels itself to have arrived, yet does not recognise this as a dead end. Importantly, virtues are not the aim of this process; they are its means. Thus, the Overman does not cultivate courage, for example, because courage is inherently virtuous or because it produces virtuous action, but because courage is integral to the Overman's creative activity.

By contrast, the Last Man – who has succumbed to the comfort of 'having arrived' – exhibits no virtues in the relevant Nietzschean sense. If one is no longer becoming, then one's virtues have crystallised and therefore cease to be virtues, for virtues are defined by their dynamic responsiveness to one's drives and

environment. Such crystallisation is equivalent to mediocrity, signalling a cessation of striving for excellence, which, for Nietzsche – as Swanton notes – constitutes “the greatest ugliness” (2005, 183). For instance, if one is always completely honest regardless of circumstances, as Kant’s categorical imperative would require, one is no longer virtuous in the Nietzschean sense, because one has ceased to be unique, conscientious, and courageous – qualities that loving fate (a concept with dynamic significance; see 5.2.1) entails. Nietzschean virtues are therefore concerned with avoiding mediocrity (Swanton 2005, 187) and fostering creativity.

Let us now consider the creativity of the Overman, which lies at the core of their self-cultivation. Nietzsche emphasises the importance of first liberating oneself from the conventions and customs imposed upon the self (Franco 2018, 54–57). Only once this ‘noise’ of inherited convention has been cleared are the conditions for sincerity and authenticity in one’s *amor fati* in place. Franco puts the point as follows: “to overcome this inauthentic disjunction between inner subjectivity and outer appearance, the honesty (*Ehrlichkeit*) and ‘truthful character [*wahrhafter Charakter*]’ of the individual must ‘rebel against a state of things in which he only repeats what he has heard, learns what is already known, [and] imitates what already exists’” (Franco 2018, 55, quoting Nietzsche’s HL 10, 123). The demand reaches beyond inherited convention to one’s own ideals, as Nietzsche puts it in *Human, All Too Human*: “we have to become traitors, be unfaithful, again and again abandon our ideals” (quoted in Franco 2018, 60, original emphasis).

Importantly, this rejection of *what is already known* refers not only to externally imposed conventions but also to the notion that self-creation is not a matter of identifying the self as a determinate, fixed entity. Uniqueness does not consist in discovering latent features of the self: “It is mythology to believe that we will find our authentic selves [...]. Rather, making ourselves, shaping a form out of all the elements – that is the task” (*KSA* 9:7[213], 361, quoted in Franco 2018, 62, emphasis added). Thus, the very concept of creativity entails rejecting discovery, predictability, and becoming *what is expected*; that is, fulfilling a predetermined role – such as the Schopenhauerian “good citizen, scholar, or statesman” (Franco 2018, 56) – in favour of generating new meaning, which presupposes the creation of new forms, departing from established patterns, and emerging from the particular experiences of the creator. In brief, the creator does not produce what is expected; they produce the unforeseeable.

This creativity manifests through engagement with obstacles and the devising of strategies to overcome them – that is, by playing the game of resistances (see 1.3). It is a game in the sense that it does not involve altering the fundamental realities

of the world, but rather engaging with them and transforming their appearances. As Miyasaki observes, “The [Nietzschean] artist does not truly create or mould the world, only its surface” (2022a, 2). It is essential to distinguish here between engaging with appearances and producing illusions. Nietzsche, in his later work, rejects art conceived as the production of illusions, since such illusions serve to conceal reality rather than confront it (Miyasaki 2022a, 2). For Nietzsche, confrontation with reality is central to self-overcoming; only through such engagement can one genuinely encounter opportunities to surmount obstacles. According to Miyasaki, this form of art is non-fictionalist and superficial in that it does not aim to transform reality, but to apprehend it in the spirit of *amor fati* (2022a, 2 fn. 2).

The Overman remains faithful to what is and employs this fidelity in the task of self-creation, while simultaneously affirming a reality in which the self is neither a mirror of truth nor its realisation. Reality simply exists, continually offering new opportunities for engagement. In this spirit, Nietzsche characterises creation as “a mocking, light, fleeting, divinely untroubled, divinely artificial” activity (Nietzsche 1888, 11.2); one that requires “above all... cheerfulness, any cheerfulness” (ibid.), and one in which “we learn now to forget well, and to be good at not knowing, as artists” (ibid.). Art, understood in this way, is a playful and continuous process of creation and re-creation, requiring the capacity to forget both to generate something new and to avoid being constrained by external conventions or by the self’s ingrained habits.

In summary, Nietzsche’s Overman is an artist – a creative being perpetually in a state of becoming. This figure does not discover meaning but creates it, generating new forms and adopting shifting perspectives, rather than merely accepting or rejecting phenomena as they appear. The rejection and affirmation that Nietzsche repudiates also involve the construction of illusions, i.e., inventing or concealing elements in order to fabricate a fantasy, thereby becoming enslaved to it (and potentially subjugating others to this slavery through the enforcement of morality to uphold such a fantasy). By contrast, the Nietzschean artist does not calculate; they create. They do not rationalise or act strategically in pursuit of a predetermined goal, selecting only the black or white stepping stones laid before them. Instead, they choose their colours and stones in accordance with light, their imagination, and their abilities at any given moment. For such an artist, life is poetry, not (Aristotelian) *phronēsis*.

We may therefore conclude our discussion of Nietzsche’s virtue ethics as an ethics of creativity with the following passage from *The Gay Science*, which captures particularly well what it means to be the poet of one’s own artistry:

“What one should learn from artists.—How can we make things beautiful, attractive, and desirable for us when they are not? And I rather think that in themselves they never are. Here we could learn something from physicians, when for example they dilute what is bitter or add wine and sugar to a mixture—but even more from artists who are really continually trying to bring off such inventions and feats. Moving away from things until there is a good deal that one no longer sees and there is much that our eye has to add if we are still to see them at all; or seeing things around a corner and as cut out and framed; or to place them so that they partially conceal each other and grant us only glimpses of architectural perspectives; or looking at them through tinted glass or in the light of the sunset; or giving them a surface and skin that is not fully transparent—all this we should learn from artists while being wiser than they are in other matters. For with them this subtle power usually comes to an end where art ends and life begins; but we want to be the poets of our life—first of all in the smallest, most everyday matters.”
(Nietzsche [1882] 1974, 299)

5.3 CONTEMPORARY ATTEMPTS TO BRIDGE VIRTUE ETHICS AND POLITICAL THEORY

Thus far in this study, virtue has been discussed primarily at the level of the individual. We now turn from the individual to the social sphere. Virtue ethics has been criticised for its inability to move beyond individual character in order to provide a framework suitable for structuring society or for theorising political concepts such as democracy, social justice, and institutional arrangements. On this view, virtue ethics is regarded as useful only for evaluating the character of moral agents, thereby confining it to moral philosophy rather than political philosophy.

However, I argue that the problem lies elsewhere than in the personal-public divide. Rather, the existing accounts, as discussed below, attempt to use the same structure of virtue ethics in the public sphere as they do in the private sphere, assuming that the public sphere requires a universal and obligation-driven conception of ethical thought. In short, the ethical accounts discussed by political theorists are approached in such a way that the virtuous agent is understood as someone who realises the political good. But this creates a difficulty, since the assumption of universality runs against the particularism that makes virtue ethics distinctive. As McDowell argues, the knowledge of the virtuous person is a capacity to see what a situation calls for, and it cannot be reduced to a universal

formula without ceasing to be the capacity that recommended it (McDowell 1979, 336).

The accounts that bridge virtue ethics and politics typically build on eudaimonist, (neo-)Aristotelian accounts and consider virtues as constitutive of human flourishing (MacIntyre [1981] 2007; Hursthouse 1999), though they diverge significantly in their substantive content. For example, Alasdair MacIntyre's ([1981] 2007) communitarianism locates the value of virtues within the social practices of particular communities rather than in abstract or universal principles. Thus, on this view, goods are internal to a practice and not specified before the practice is entered. In this respect, MacIntyre seems to come close to the order of reasoning defended in chapter 6, but this would be a misreading, as MacIntyre's virtues are the means by which an inherited good is realised – and here the inheritance, not the individual, does the determining. By contrast, varieties of neo-Aristotelian liberalism address the question of the state's role as a shaper, guarantor, or protector of flourishing (Rogers and Russell 2025, 368–371). Here, the central challenge lies in determining what counts as flourishing and how authority and legitimacy factor into the imposition of a particular interpretation of the good life, thereby reproducing the familiar liberal dilemma between negative and positive liberty, while presupposing consensus and cooperation.

In this way, when the focus is on substantive political outcomes – that is, on the particular ends political order is meant to realise – virtues are used to construct an ethical legitimacy of political authority, and ethical concepts are shaped to align with political principles, for example by serving to justify particular models of state authority (Rogers and Russell 2025, 371–374). Because this approach resembles the way moral principles and norms are used in political theory, i.e., as universal and obligation-producing, we can describe such a merger of virtue ethics and political theory as *virtue moralism*.

To illustrate this, let us consider Mark LeBar's (2009, 2013) account of how virtues provide reasons for justifying political authority that entails obligatory constraints. LeBar argues that the principle of equality governing interpersonal relations extends to the relationship between individuals and state institutions. Since the reason for acquiring the virtues – namely, the achievement of human flourishing – is the same reason for endorsing state regulation of our shared social life, it follows that delegating authority to institutions and accepting obligations to comply with their regulations is justified, insofar as the flourishing of the individual depends upon the flourishing of others. Institutional authority is thus treated as continuous with the equality characteristic of interpersonal relations. Rogers and Russell argue that this approach risks depoliticising politics, as the

state is, after all, by design required to have greater authority than individual citizens, and if this difference is denied, the role and purpose of the state become unclear (2025, 372).

As another example, Onora O'Neill's (1996) Kantian account merges virtue and justice, which unifies a universalist theory of justice with an account of the virtues through a shared standard of practical reasoning. Moral requirements are articulated as obligations encompassing both justice and virtue, rather than as ideals of character detached from action. Virtues matter, on this view, because they enable agents to reason and act justly. Justice, for O'Neill, structures relations among agents and cannot be treated as a purely abstract or self-implementing ideal. This leads her to reject highly idealised theories of justice, such as Rawls's, in favour of a relational account in which justice is constituted through agents' interactions rather than imposed from an external moral standpoint.

O'Neill's own proposal argues that because agents are limited in knowledge and capacity, principles of justice must be worked out collectively and aim at protecting agency itself rather than an idealised moral order. As O'Neill puts it, where assumptions about the political sphere do not presuppose "conjunction of forms of plurality, connection and finitude, the scope of ethical standing and consideration will remain wholly undefined" (1996, 101). On this basis, she distinguishes obligations of justice, arising from the rejection of injury, from obligations of virtue, arising from the rejection of indifference and neglect (*ibid.*, 205). On her account, both are indispensable, since injury undermines agents' capacity to meet the demands of justice (*ibid.*, 164), thereby generating duties of social virtue, such as solidarity and rescue towards others (*ibid.*, 197).

From these examples, we can see that introducing obligation into virtue ethics changes how virtue is understood. Once a particular goal or good is established in advance, virtues then become means for achieving it, making the rationale of ethical thought structurally similar to deontological and consequentialist theories, which likewise begin with some principle or desired outcome and then use it to determine how people ought to act.

This can also be seen in Judith Shklar's discussion in *Ordinary Vices*, particularly in the chapter "Bad Characters for Good Liberals". Shklar treats virtues and vices partly in terms of how they affect moral action (1984, 232-236), thereby placing the focus on action, with the character of the agent becoming relevant to how that action is understood. Consequently, a good person can do something morally wrong, while a bad person can sometimes do a morally right thing. Similarly, good people may sometimes have to dirty their hands (see [Coady 2024](#)) by doing something morally wrong in order to achieve a morally good outcome. Thus, the

main concern is still what counts as the ‘right’ action, and virtues are assessed according to whether they help the agent arrive at that action.

A shared conception of the good therefore creates a difficulty for virtue ethics, because it creates obligations, turning virtues into general requirements that apply to all. This can obscure the distinction between virtue ethics and principle-centred approaches, which focus on determining the right action. Consider Hursthouse’s (1999) account. Her concern is with applying virtue-based reasoning to particular situations rather than developing a comprehensive political theory or examining the broader question of how political problems should be understood (Swanton 2021, 4). However, any theory that tells us what to do must first assume that the problem has already been identified and framed in a particular way, but since political problems are defined in terms of structures, interests, or institutions, while virtue ethics conceptualises the situation in terms of what it asks of a particular agent in the light of the good, a problem defined in the first way may not give virtue anything meaningful to respond to. Thus, such action guidance, otherwise entirely reasonable in itself, may still be directed at the wrong problem.

5.4 TO CONCLUDE: ON THE IMMORALISM OF VIRTUES

The aim of this chapter has been to examine Aristotle’s, Slote’s, and Swanton’s accounts of virtue ethics, together with that of Nietzsche, who is not usually discussed in the same context, but nevertheless expresses an ethical thought rejecting morality and who has influenced a number of anti-moralist thinkers, including this study. Each of the first three accounts considers virtue to realise value, although they do so in different ways, with Aristotle locating it in the activity, Slote in the motive, and Swanton in the target to which the virtue responds, and all of these consider virtues as a response to a situation. Nietzsche, on the other hand, can be interpreted as considering virtue as a proactive, productively disruptive attitude. This is why the chapter ends with Nietzsche and the next chapter builds on his conception of ethics.

Virtue ethicists disagree about what virtues do. For Aristotle, virtues are stable states of character in accordance with which a person acts well. For Slote, virtues are the motives from which right action arises. For Swanton, they are dispositions that enable a person to respond well to the demands of a particular field. But all three share the idea that the good is specified independently of the exercise of virtue, and that the exercise is then assessed against the standard the good sets – virtue works insofar as it contributes to the good.

Virtue ethics places the person, rather than the principle of right action, at the centre. In this sense, the difference between morality and ethics can be expressed as a shift from asking what we owe to one another to asking how we are towards one another.

Nietzsche's anti-moralism rejects the idea that the good can simply be discovered as something objective, fixed, or predictable. For him, values do not have an inherent meaning. Their significance develops through the ongoing interaction between the self and its environment, in other words, through experience. Creativity, both in the creation of the self and in the creation of values, is therefore central to human existence. Values arise because human beings value things; valuation is an active process. Human beings are therefore not simply discoverers of meaning but creators of it. Meaning is something achieved through aesthetic and existential activity rather than something objectively present in the world waiting to be discovered.

Nietzsche's ethics is therefore concerned with becoming rather than being. In this character-based approach, virtues are dynamic aspects of the self that contribute to the ongoing process of becoming through engagement with obstacles. Nietzschean virtues are conditions for self-creation rather than universal traits that guarantee morally correct behaviour.

For Nietzsche, creativity requires moving beyond what is familiar (and thereby comfortable) – social norms, inherited ideals, and even previous versions of oneself – to produce something new and unforeseen rather than simply fulfilling predictable expectations. This is what I mean when I argue that virtues cease to be virtues once they become crystallised (5.2.2). It is also why the account developed in chapter 6 understands virtues as capacities rather than stable traits.

In 5.3, I introduced the concept of *virtue moralism* to designate the merger of virtues as character traits with moral obligation, for instance, when virtues are used to justify a political order, as in LeBar's derivation of institutional authority from the equality of interpersonal relations, or in O'Neill's account of virtue as an obligation to reject indifference and neglect, thereby creating duties of social virtue, such as solidarity and rescue, towards others. In this way, virtue ethics can reproduce the same structure as the approaches it was intended to replace, since virtues are still understood as realising some preset value.

With this in mind, we can now turn to the final chapter, which examines how normativity in politics can be understood as a virtue-ethical project.

6. THE ETHOS OF CREATIVE VALUATION

6.1 PRELIMINARY OUTSET: VIRTUES, DISPOSITIONS, AND VALUE

We have now reached the final chapter, towards which the preceding discussion has taken us. Here I lay out the proposition of how virtue, and in particular the virtue of creativity, can guide theorising about politics.

In the first chapter (1.4.2), I discussed in detail the conception of ethics I adopt in this study. Let us briefly recall: ethics is an epistemic enterprise concerned with how values acquire meaning through agents' interpretive and creative engagement with the world, and with human beings as the agents of that valuation. Ethics deals with relevance, approached through an attempt to understand how values acquire their meaning and what that meaning signifies. I also argued that ethical thought would be empty if it did not include human character traits, for the reason that, for ethics, persons are the object of inquiry, not their actions, and thus, any adequate account of ethics will, in the end, become an account of the virtues.

“Virtues are values related to human character” – this is how Brobjer (2025, 29) elegantly puts it in his paper on Nietzsche's revaluation of values. Here, Brobjer helps us set out with two points of departure. The first connects virtues to the creation of value, and the second brings us to Nietzsche, on whose conceptions of morality and ethics this study draws heavily, though not exclusively, as discussed in 5.2. As Brobjer observes, Nietzsche is oriented towards character rather than abstract values, speaking less of ideals than of a new conception of the human being (*ibid.*), which reinforces my argument that ethical valuation concerns the kind of being a person is, rather than simply the rightness or wrongness of their actions.

While Nietzsche not only inspires but also underlies the ethos of my proposed creativity-centred virtue ethics, I extend his conception from the realm of the individual to that of social relations and, subsequently, political arrangements. I focus not only on Nietzsche's question of ‘what one can become’ but also on the question of ‘how we can be towards each other’.

Since this chapter aims to outline a framework for virtue-ethical theorising in politics rather than to provide a comprehensive survey of virtue ethics, areas such as metaphysical debates, psychological accounts of virtue, and empirically oriented discussions are largely set aside here. The aim of this study is to recover the theoretical significance of ethics for political thought and discuss how the

epistemic function of virtue can help reconceive the sense of justice beyond the deontological-contractual and consequentialist-utilitarian paradigms that continue to dominate modern political thought, by also incorporating the sensibility of justice. I use sense and sensibility to distinguish two dimensions of cognition, where sense concerns the (primarily rational) intelligibility of what something means, while sensibility concerns how its significance is felt and perceived.

Chapters 2–4 examined moralism, realism, and agonism as styles of normative theorising about politics, each offering its own account of the aim of politics, while chapter 5 surveyed virtue ethics. The argument has moved progressively closer to the position developed in this chapter. My diagnosis is that the three political perspectives examined converge in *prioritising value over valuation*, which means that they take what counts as valuable (and the meaning of value along with it) to be determined prior to the act of valuation. This is why the structure of moral reasoning has served as the primary framework for examining what role values should play in politics.

This work argues that an alternative can be found in ethical valuation. Agonists have gestured towards such an alternative, as has Williams (1985) from the moderate realist camp. Yet neither has developed this alternative systematically. One reason is that ethics has traditionally been understood as belonging to the private sphere and therefore as unsuitable for public political theorising, insofar as it lacks the universalising ambition that morality readily provides.

My proposal, therefore, is an ethics-first approach to political theorising that is not moralising, as it does not seek universal principles, determinate obligations, or definitive solutions. Instead, the aim is to offer a values-first alternative to approaches that either affirm morality as the ground of political normativity, as moralism does, or reject it altogether, as realism and agonism tend to do, while nevertheless retaining values at the centre of political theorising. To this end, chapter 5 outlined the core features of virtue ethics. We are now in a position to examine the possibility of placing virtue ethics at the centre of political theory. In doing so, the analysis is guided by the conceptions of ethics and morality developed in section 1.4.

6.2 WORKING DEFINITIONS: VIRTUE, DISPOSITION, AND GOOD

Before setting out the framework, it is useful to emphasise the way the conception of virtues in this study departs from the more common understanding. Virtue ethics usually assumes that virtues help us to discover, understand, or achieve the

good. For example, virtue consequentialists, such as Driver, define virtue as “a character trait that produces more good (in the actual world) than not systematically” (Driver 2001, 82), and Hurka’s virtue consequentialism holds that virtues are valuable in themselves because they express the right attitudes towards things that are already intrinsically valuable (2001, 20), while Swanton’s pluralist virtues are dispositions to respond well to items within a field, where the field and its demands are given in advance of the response (2003, 20). Some virtue ethicists reject this stronger view. For instance, Hursthouse (1999) argues that we partly discover what a good life is by living virtuously. But this is only an epistemic claim, as Hursthouse still assumes that there is an underlying standard of human flourishing, grounded in facts about the kind of creatures we are, which the virtues help us discover.

My proposal reverses this assumption, arguing that the good is not something that exists independently and is discovered or pursued through virtue, but that it is constituted in the exercise of virtue itself. On my view, virtues are capacities that have an *epistemic* function; they create dispositions through valuation, and it is through them that (social) values acquire their meaning.

As a running illustration throughout this chapter, I will use Bernard Williams’s (2002) account of truthfulness, which consists of two virtues, namely accuracy and sincerity, which underlie the social value of trust. Accuracy involves doing what one can to acquire true beliefs, while sincerity involves ensuring that what one says reflects what one actually believes (*ibid.*, 11).

6.2.1 Virtue as a capacity

By virtue, I mean a narrow character trait or form of functional excellence that constitutes a capacity for ethical valuation. On its own, such a trait has no evaluative meaning. For example, one may be accurate or sincere, either consistently or occasionally, without either quality being ethically significant. A virtue by itself does not tell us what matters in a given situation; it is merely a capacity that remains dormant until it is directed towards something.

Virtues have an epistemic role, in that they are ways of valuing things, through which values acquire their meaning and become the ‘good’. So, on my view, the good is not something that simply ‘is’; rather, it is a value whose relevance, worth, and substance lie in the meaning that the virtuous agent attributes to it.

Virtue, as a trait, is therefore not self-standing. It might be tempting to consider virtue similar to a moral principle, but they actually work in very different ways. A principle such as ‘lying is wrong’ already tells us what is considered important and gives us a standard for judging an action before we encounter a particular

situation. A virtue such as honesty does not work this way. On its own, honesty does not tell us what matters or what we should do. It is a capacity that becomes active through attention and is given meaning through the process of valuation, which directs how that capacity responds to a particular situation. In other words, honesty can have different content in different situations for the same agent, because what it means to be honest depends on many other considerations relevant to the particular situation at hand.

Some virtue theorists distinguish between intellectual and moral virtues. On this view, moral virtues, such as respect and empathy, concern interpersonal relationships and promote well-being, justice, and other moral goods, while intellectual virtues, such as open-mindedness and curiosity, focus on the pursuit of knowledge and truth and enhance our cognitive capacities (Turri et al. 2021).

However, this study draws no distinction between intellectual and moral virtues¹⁵, considering all kinds of virtues relevant to *ethical valuation* as virtues in my sense (the distinction between ethical and other kinds of valuation is discussed in 6.4), because I view the two as overlapping in practice. The example of truthfulness illustrates this well, since accuracy is an intellectual excellence, while sincerity is an ethical one, and a division between the two would therefore divide a single disposition into two. This notion also follows from what I argued in 1.4.2 about affective cognition: if the apprehension of what matters is affective throughout, rather than a rational operation with feeling added to it, then virtue cannot belong entirely to one side of the division.

The inseparability of moral and intellectual virtues is also supported by the responsibilist camp of virtue epistemology scholarship (Montmarquet 1993; Zagzebski 1996; Baehr 2015), which acknowledges the interdependency between intellectual and moral virtues, with the former considered rational virtues and the latter subjective and sometimes affective ones. Zagzebski notes, for example, that honesty as a moral virtue requires diligence in seeking and communicating truth (1996, 158), and that several virtues, such as courage and humility, overlap in their moral and intellectual dimensions (ibid., 159).

Further, I do not take virtues as features of action-guiding principles grounded in their ontological properties, as Julia Driver (2003) and Alan Wilson (2017) suggest. I consider them instead as capacities that do not guide action directly, but shape how agents interpret, evaluate, and make sense of normative considerations. Driver and Wilson argue that treating moral and intellectual virtues as the same kind of thing makes it harder to see the conceptual difference between them. I

¹⁵ I retain this term ‘moral virtues’ here to align with general convention, although my view does not discuss moral reasoning as such.

agree that this would be a problem if these different properties were supposed to perform the same reasoning function. On my account, however, the important distinction is between a capacity and the disposition through which that capacity is expressed. The difference is therefore not a difference in subject matter, but in what each does in the process of evaluation.

This study also departs from the conception of virtues as skills proposed by Julia Annas, who compares them to skating or playing the piano, and holds that one may practise without performing perfectly (2011, 65). While I can accept that virtues are “an essentially developmental notion” (ibid., 38) and that “we do not have an adequate account of a virtue until we have an account of how it is taught and learned” (ibid., 163), I resist the analogy itself, because I take a skill to have a standard of success that lies outside its exercise: a piece of music can be played well or poorly, and such a standard does not depend upon the disposition of the player. This illustrates the priority of value over valuation, and it is the same difficulty I identified in Swanton’s fields (5.1.3) and in Aristotle’s function argument (5.1.1). On my account, a virtue has no such external standard, for the reason that it has no evaluative content until it is taken up into a disposition, and so it cannot be more or less well executed in the way a skill can. Conceptually, then, I distinguish virtue as a raw trait, which I understand as a capacity, from a virtuous response, which I understand as a disposition; and neither should be understood as a skill.

6.2.2 Disposition

By *disposition*, I mean a way of seeing and evaluating things that is shaped by virtues, rather than simply a habit or tendency to behave in a particular way. A disposition develops when a virtue, understood as a capacity, is directed towards something. The capacity itself is dormant until it is brought to bear on a particular object or situation. In this sense, a disposition develops when a virtue transcends the abstract state of capacity and becomes a way of responding to the world more generally.

In short, a virtue is exercised by valuing a particular situation, and the considerations that matter in that situation shape what honesty means. Over time, as the agent encounters more situations that call for honesty, they develop a disposition to respond to such situations in certain ways, yet each new encounter also changes what honesty means to them and how that disposition shapes their way of responding to the world. Put otherwise, throughout life, the agent continues to think of themselves as honest, but what this means varies, since the justification for considering oneself honest changes in each particular

circumstance, even if only slightly. Thus, the capacity of a virtue acquires its meaning through valuation, that is, through its exercise.

Now consider the virtue of sincerity and the disposition of truthfulness. Sincerity initially concerns what I convey to others, i.e., whether what I express is what I actually believe. Truthfulness develops when this concern with sincerity extends to other aspects of how I relate to myself and the world, transparently and reflexively. Accuracy becomes part of this because I cannot be truthful about my beliefs without making a genuine effort to get things right. This is why Williams treats sincerity and accuracy as two aspects of truthfulness rather than as separate virtues (2002, 11). Yet neither sincerity nor accuracy, even when coupled together, is evaluative in itself. They are still capacities that become evaluatively meaningful only when they are directed towards something.

Truthfulness as a disposition is therefore a way of relating to something that can be evaluated, and it is through such relationships that things come to matter in a particular way. For example, a truthful person can consider whether someone is being deceptive or evasive, whereas, for a person who is not truthful, the concept of truthfulness may remain alien or obscure. If sincerity involves making an “assertion [that] expresses what one actually believes” (Williams 2002, 96), then without sincerity, an evasive or untruthful response may instead appear simply as confusion.

This means that judgement about truthfulness can go wrong not because a person applies the wrong virtue, but because they fail to recognise what is there to be judged in the first place. An untruthful person may lack, or suppress, the relevant experience and therefore fail to recognise something as an object of evaluation. The relationship therefore works in both directions: a virtue becomes a disposition when it is directed towards something, but the virtue itself is also necessary for that disposition to develop and for the corresponding good to emerge.

6.2.3 Good and bad virtues?

While virtue-ethical accounts differ considerably in how they understand the relationship between character, action, and value, as we also saw in chapter 5, they tend to converge in considering virtues and vices as distinct in evaluative terms, where vices (e.g., meanness) are the evil counterparts or mirror images of virtues (e.g., kindness). However, if we understand virtues as involving many different concerns and targets, as Swanton (2003, 2021) does, it becomes harder to differentiate character traits straightforwardly in terms of their goodness or badness. Only a few recent works have begun to discuss whether vices form a unified category at all (Kupfer 2011; Azzano and Raimondi 2023; Robson 2024).

The following discussion likewise argues that this simple polarity between virtue and vice is insufficient in light of the account developed in this study. Thus, although a full discussion is beyond the scope of this study, it is nevertheless useful to consider the distinction briefly, as doing so makes clearer how virtues give meaning to values by shaping dispositions.

Scholarship in virtue ethics has focused on identifying which traits count as virtuous and which as vicious, often in relation to competing conceptions of the good life. For example, Müller observes that “[a]lthough the vicious agent has the wrong conception of the human good (happiness), she still organises her life around this conception just as the virtuous person does so around the correct one” (2015, 459). In other words, for Müller, a vicious person has the wrong conception of what is good, thereby presupposing that there is such a thing as a correct conception of the good.

The alternative developed in this study rejects the idea that there is an already established, objectively correct conception of the good against which particular valuations can be judged as right or wrong. On this view, the problem is that if we tried to use a fixed idea of the good to judge particular valuations (as the concept of virtue and vice entails), that idea would itself require justification. This means that it would have to be derived from some earlier valuation, after which we would have to ask why that earlier valuation is relevant to the case at hand. We would therefore have to appeal to some foundational value if we wanted to avoid an infinite regress. Within the perspective proposed in this study, however, such a foundational value is problematic because valuation is understood as situated and relational, since, as I argue in section 6.3.1, even apparently basic notions such as moral equality do not have a uniform meaning when considered independently of the circumstances in which they are applied.

On my account, then, valuation is not about determining what people ought to value, but rather about how valuation should take place. This is not to imply that there can be no criteria for good valuation or good disposition-creation. Rather, a fuller account of such criteria must be left for future work, although some initial criteria are discussed in sections 1.4.2 and 6.3.2.

Throughout this study, I therefore continue to employ the term ‘virtue’, but I use it in an evaluatively neutral sense. Virtue, as such, indicates neither goodness nor badness. Rather, it refers to a capacity that contributes to the creation of dispositions, through which agents create the meaning of social value. To be sure, I do not mean to reject the judgement that some dispositions may result in harm and others in joy, or that harmful effects should be disregarded. This simply means that, on the view discussed here, valuation cannot begin from virtues if those already have some substantive value (vicious, virtuous) attached to them.

Thus, such a distinction does not work in the context of this study, because it prioritises the process of valuation over the assumption that the substantive content of virtues ought to be fixed in advance.

6.2.4 Social value

By social value, I mean a relational good that emerges when a disposition is exercised and taken up by others. Trust is the social value that emerges from truthfulness, although it should be noted that truthfulness is a necessary but not sufficient condition for trust. The move from disposition to social value is not quite the same as the move from virtue to disposition, because virtues form part of dispositions as capacities. Truthfulness operates within an agent, shaping what they notice and pay attention to and how they judge, whereas trust exists only between agents. For trust as a good to exist, therefore, both parties have to recognise the disposition of truthfulness and respond to it. This is also where the question of acquisition returns, since a disposition is formed and confirmed as a disposition, rather than a passing unilateral impulse, precisely through such relational exercise. This is also how Aristotle's observation that "it is from acting as we do in our transactions with other human beings that some of us become just and others unjust" (NE II 1 1103b14–15) survives in my account, even though his teleology does not.

On a more conventional account of virtue ethics, truthfulness would be understood as a value that exists prior to the virtue, while honesty, sincerity, accuracy, and related traits would be understood as ways of expressing or realising that value. The development of virtue would then be explained through habituation (Aristotle), the cultivation of the right motivation (Slote), target-adequacy (Swanton), or some similar account. The scholarly focus would therefore be on how, for example, honesty realises or makes possible an already established good of truthfulness.

By contrast, on my account, trust is the result of an exercised disposition of truthfulness, that is, a disposition formed through virtues such as honesty, sincerity, and accuracy. Of course, different configurations of these virtues may yield different understandings of the value of trust.

Sincerity makes trust possible by requiring an absence of deception, while still allowing someone to be genuinely mistaken. Accuracy-based trust, by contrast, depends on care and diligence and is undermined by negligence. Honesty, in the Nietzschean sense I use in section 5.2.2, means refusing to hold on to comfortable beliefs, especially about oneself. Trust based on honesty is therefore trust in someone's willingness to have their flattering beliefs challenged and corrected,

and it is undermined when someone maintains a flattering view of themselves that neither they nor others are willing to question.

Thus, on my account, virtue is not a starting point or a means for achieving a value that has already been decided. This is the case whether that value is understood as an activity (Aristotle) or as an end-state (virtue consequentialism). Instead, virtue is a capacity that becomes a disposition when a situation calls for it. This does not mean that our capacities are fixed or simply given to us. On the contrary, they are also shaped by how we relate to the world. Our relationships with others and with our surroundings can make it easier or harder to develop and exercise a capacity, and more radically, social interaction can even create capacities that did not exist before.

6.3 FROM VIRTUE TO DISPOSITION TO SOCIAL VALUE

6.3.1 Virtues as epistemic capacities for valuation

My virtue-ethical proposal in this study begins from the rationale that there are different ways of understanding the diversity we encounter in society. Chapters 2–4 showed that moralism, realism, and agonism tend to understand diversity primarily as a diversity of commitments, including foundational commitments that are taken to be irreconcilable. Structural diversity is also acknowledged, but generally plays a secondary role.

This matters because what appears as the same commitment may arise from very different underlying reasons. Consider, for example, the belief that abortion is wrong. One person may hold this belief because they believe that human life begins at conception; another may hold it because they believe that God has forbidden abortion. Both may hold the same commitment with equal strength, yet the reasons underlying it are fundamentally different.

These reasons, moreover, are not themselves indivisible entities. They are composed of different elements, including values, and the values we hold constitute a significant part of our commitments. This work therefore understands social diversity as a *diversity of reasons*, which entails a diversity in the *meanings* that *values* acquire and, consequently, different conceptions of what a given *problem* consists in.

From this perspective, the difference between the approaches examined in chapters 2–4 and the virtue-ethical approach proposed here now becomes clearer. Moralism, realism, and agonism are primarily concerned with how the social sphere should be organised – whether through agreement, legitimate stability, or

perpetual contestation – and thus focus, to varying degrees, on the regulation and coordination of behaviour. A virtue-ethical approach, by contrast, begins with the reasons underlying behaviour. It addresses these reasons through the process of valuation, by asking not simply what people value, but *how values acquire meaning* through the dispositions and practices of those who enact them.

Understanding reasons behind commitments is, of course, not a straightforward endeavour. But our task here, in political philosophy, is not to seek explanations for observable phenomena (reasons as causes) but to explicate the schemata for normative theorising. That is, reasons feature here as something a theorist is required to prioritise over outcomes or principles or justification of particular values. Reasons are not, however, the same as justification, because to justify a conception of trust is to take it as already determined and to argue that others should accept it, but the reasons I am concerned with come earlier: they are what makes a conception the conception it is, rather than what backs it once it has been formed. This is why justification, however scrupulous, cannot answer the question I am asking, since it begins where my question has already been settled.

So, when virtues, as capacities that form the dispositions which give social values their meaning, are used as a methodological outline, we are directed to focus on constitutive reasons rather than instrumental or pragmatic ones. Consider again the example of truthfulness. If we were to start arguing that trust is the gold standard of democratic institutions, then without understanding what this trust entails, this statement would be hollow and empty, and all roads would be open to substantiate the conceptions of trust as we desire and can justify. And this is, of course, exactly what substantive normative theories do; this is the proper business of normative theorising, considering what would be a substantively good society, whether as ideal or non-ideal theorising.

But my project takes a step back and asks a different question: what makes us value certain kinds of social relationships in the first place? The answer may initially seem obvious. Trust, for example, might be considered valuable because it enables cooperation and thereby contributes to prosperity and well-being. The more difficult question, however, is not whether trust is generally good for social relations, but what exactly we mean by trust and what makes it valuable. This is where normative theorising becomes important, since rather than simply assuming that trust is valuable, it asks why trust matters in the first place and what, precisely, makes it worth valuing.

The same applies to widely agreed-upon concepts of autonomy, moral equality, care, avoidance of harm, and so on. It would not be wrong to say that, at face value, these concepts can be supported by diametrically different reasons and

substantive content. These concepts are therefore placeholders whose meaning is filled in through social practice.

So, if we agree that my project of uncovering the reasons behind certain conceptions of values (as ethical goods) has traction, why should we choose ethical virtues and dispositionality for this? There are two reasons. First, I follow an underlying foundational commitment to the notion that ethical thought underlies other considerations in the social realm. Recall from 1.4.1 that the guiding question that distinguishes morality from ethics, as I conceptualise them, is ‘what we owe to each other’ vs. ‘how we are towards each other’ – this, in my view, takes priority over any instrumental considerations, e.g., that humans are more efficient when they cooperate peacefully. Second, if we agree on the priority of the ethical, the next choice is whether to focus on substantive ends, metaethical grounding, or epistemics. If the question is about understanding how values acquire meaning, and thus the reasons behind certain valuation choices, we need to understand the process through which that meaning comes about. The question therefore becomes: ‘How do we know trust is valuable?’

This brings us to yet another choice, which is also a pre-theoretical commitment. If we hold formal, concept-coherence-driven theorising about principles and rationality in high regard, we might choose moral reasoning as the object of inquiry. But if we consider affective cognition or lived experience to be part of ethical thought, the messiness of reasoning becomes difficult to accommodate within formal theories of morality that seek solutions and right action. We can instead focus on the human being, with their capacities for valuation and dispositionality as means of meaning-making.

In sum, the virtue-disposition-value chain matters because it directs our attention to how meaning is made. Moralism, realism, and agonism theories take trust as a social value and ask what secures it, whether as a guarantor of cooperation, stability, or agonistic respect, and traditional virtue theories tend to focus on how virtues can be cultivated or developed so that they produce trust. My question asks instead not how trust is secured, nor how it is produced, but what conception of trust is at stake in a given case, and which capacities and dispositions give it that particular meaning.

6.3.2 Virtues and the process of valuation

Before tracing the process of valuation itself, I should first establish what distinguishes ethical valuation from valuation of other kinds, such as aesthetic or pragmatic valuation. I have already outlined the difference between ethical valuation and ethical reasoning in 1.4.2, and what follows builds on that distinction.

I consider four features of ethical valuation as most relevant for this study. First, it is *relational*, concerning how we are towards one another (see 1.4.2), and therefore bears on the relationship between the person doing the valuing and the object of valuation, as well as on the context in which that relationship stands. Second, it is *reflexive*, meaning that the valuer must be answerable to the same value that they create. Reflexivity is crucial for ethical valuation because it means that whoever creates or upholds a value must also be willing to apply it to themselves. Third, it carries an *affective* dimension and is directed towards the character of what is valued (Goldie 2000, 151–160), which prevents ethical valuation from being reduced to a purely perceptual process. What Goldie calls “feeling towards” (2000, 51) is not an emotion arising once we have already perceived something; our feelings and dispositions are part of how we perceive and understand it in the first place. The affective dimension of cognition therefore shapes what we see in a situation and what that situation means to us. Fourth, ethical valuation is *internally motivated*: the valuer does not esteem something because they follow some preset standard or rule. While an external standard can govern conduct, it cannot create the meaning of a value because it is not the valuer’s own. Although, conceptually, some virtues can be understood as rules (see the discussion of virtue moralism in 5.3), in my understanding, a disposition ceases to count as a virtue, however admirable or well exercised, when it no longer originates in the valuer’s own valuation and instead amounts to mere performance.

How, then, does virtue give rise to the dispositions through which value is created?

In 1.4.2, I discussed Iris Murdoch’s example of mother-in-law M and daughter-in-law D (Murdoch [1970] 2014, 16–24), in which valuation was grounded in attention not only to the object of valuation and its context, but also to the valuer’s own beliefs. In Murdoch’s example, M gradually comes to see D differently without receiving any new evidence, reaching a reasoned decision, or responding to external pressure or motivation. What changes is her disposition.

Viewed through the lens of virtue, we can interpret this process as beginning with M’s virtue of honesty. Honesty enables the attention through which a new perspective can emerge, and M’s willingness to examine her own beliefs critically makes it possible for her to see D differently in the first place. What makes honesty a virtue here, rather than a disposition, is that it does not determine what M must see or conclude. Instead, it enables her to examine her own judgement honestly.

The problem with M’s initial judgement was not that she failed to notice D’s behaviour, but how she understood it. She noticed the same qualities from the

beginning, but initially saw them in a negative light; then she turned the same critical attention towards herself and asked whether her own view of D might be mistaken. In doing so, she applied the same standard of honesty to her own judgement that she had applied to D. What changed, then, was not D's behaviour but how M saw it, since what she had first described as "tiresomely juvenile" she later saw as "delightfully youthful" (ibid., 17).

The example therefore shows how virtue leads to valuation, and valuation leads to disposition. M's virtue of honesty allows her to pay attention in a different way, which leads her to value D differently, and this new valuation changes her disposition towards D. In other words, disposition is not the starting point of the process but the result of valuation, through which M comes to see and respond to D in a different way.

This understanding of virtue as an ongoing formation of disposition connects with Nietzsche's account of honesty, which underlies the formation of a disposition as an ongoing process of negotiation towards a more harmonious way of being, without ever reaching a final state of complete harmony. For Nietzsche, honesty is "my virtue" ([1883] 2006, *On Science*, original emphasis): a virtue that takes a different form for each individual in their unique path of creation and self-affirmation. He calls it "the youngest of virtues" ([1883] 2006, *On the Hinterworldly*, 22) and understands it as the capacity to face difficult truths, not because a universal moral law commands us to do so, but because acknowledging such truths accords with reality and enables the creation of one's life by "tolerat[ing] a greater degree of truth" (Dennis 2019, 66 fn. 42), thereby fostering self-cultivation.

Interpreting M's valuation through a Nietzschean lens also shows us how self-cultivation takes place through creative meaning-making (for more on creativity as a virtue, see 6.4). The creation of value is an active process, in the sense that it is not merely an appropriate response to something external but emerges from within the individual. As Nietzsche says, esteeming is creation (*Schätzen ist Schaffen*): "humans gave themselves all of their good and evil. Indeed, they did not take it, they did not find it, it did not fall to them as a voice from heaven. [...] That is why they call themselves 'human,' that is: the esteemer" ([1883] 2006, *On a Thousand and One Goals*). Valuation, then, does not simply assign an existing label to a familiar or unfamiliar situation; it creates a new meaning for the esteemer. Nor does valuation terminate in a verdict, whether deontic (permissible or impermissible), axiological (good or bad), or aretaic (admirable or contemptible). Instead, valuation establishes relevance through meaning-making. Values emerge from this process; they do not exist prior to valuation in the sense of relevance-making.

Attention

This brings us to a question that has so far remained implicit: what engages the virtue in the first place? The answer is *attention*, again drawing on Murdoch. To recall, attention is Murdoch's ([1970] 2014) central concept (see 1.4.2). Although Murdoch herself does not define attention, shifting its emphasis across contexts, it can be roughly described as a task or effort towards goodness, and as a way of inquiring into moral truth. To understand this better, let us borrow Caprioglio Panizza's (2022) characterisation of Murdoch's attention as working in three registers: faculty, attitude, and activity. As a faculty, it is the coordinated use of the senses, intellect, and emotion together. As an attitude (or attentiveness), it is an orientation towards reality that, through habit, can settle into something like a virtue (in the Aristotelian sense): a trait that makes attending feel spontaneous and characteristic of the person. But attention can also be a one-off act, performed even by someone who is not generally considered an attentive person.

Such a characterisation reflects Murdoch's standing as a broadly Platonic moral realist (Blum 2023). However, Murdoch's idea of reality should not be understood as something entirely objective. Murdoch distinguishes moral knowledge from "impersonal quasi-scientific knowledge of the ordinary world" ([1970] 2014, 37). Instead, she describes knowledge of goodness as "a refined and honest perception of what is really the case" (*ibid.*). Put simply, Murdoch's moral agent sees the world through her own concepts, experiences and sensibilities, not from some impersonal standpoint, and my approach starts from the same place.

However, I do depart from Murdoch over the issue of whether the ethical value we perceive already exists before we perceive it. For Murdoch, moral reality is already there, and the task of the moral agent is to perceive it as accurately and "lovingly" (Murdoch [1970] 2014, 33) as possible, even though each person will interpret it in their own way.

On my account, by contrast, value is not simply discovered, but created through the encounter between the person and what they encounter. Consequently, attention is an internal motivation and a particular way of looking at the world. Such looking then results in *seeing* what is relevant and how to interpret it, which activates the otherwise dormant virtue. Given that virtues are themselves capacities requiring activation, this raises the further question of what activates attention. Murdoch's answer is the magnetic pull of the Good, drawing the agent through Love (Murdoch [1970] 2014, 100). Rather than seeing this process as centred on the good, I locate its impetus in the Nietzschean drive towards self-cultivation discussed above. Attention is not engaged by some external source of value, but by the agent's own desire to be honest, i.e., to tolerate more truth, and create their own life. Attention is therefore not an additional capacity behind the

virtues, but the form that the agent's own drive takes when they encounter a situation that calls for valuation.

This has two implications. First, while Murdoch's attention is discovery-oriented, in that attention reveals a moral reality (Caprioglio Panizza 2022, 159), on my account attention participates in constituting reality, since valuation does not check against an antecedent standard but creates new meaning through the encounter itself, in the Nietzschean sense of *esteeming as creation* (5.2.1). Viewed this way, attention is not in the business of correcting for distortion so as to see what was already there, but of sustaining the agent's engagement with a capacity as it generates something new. This is not to say that attention is careless about the accurate perception of facts; quite the opposite – attention entails attentiveness both to the object and to one's own possible distortions. It simply goes further, working from the circumstances as they are to constitute a new meaning for the agent. Such attention therefore implies due diligence towards the actual circumstances, so as to avoid slipping into fantasy or comforting interpretation, while the disposition that emerges from valuation reflects what comes to matter most to the valuer.

Second, unlike Murdoch's third register, where attention is an "activity, which has a beginning and end in time" (Caprioglio Panizza 2022, 158), on my account attention, once engaged, operates continuously, as valuation itself cannot be declared finished at any particular point. Attention, in this sense, is coextensive with valuation rather than merely a temporary and functional part of it. The emphasis accordingly falls on reflexivity and on attending to one's own possible illusions, rather than on uncovering a reality that precedes the encounter. I therefore retain what Murdoch terms "unselfing" ([1970] 2014, 82) while departing from the antecedence of the moral reality that unselfing, for her, discloses.

In parallel to unselfing, Murdoch's attention involves "a loving gaze" ([1970] 2014, 33), where she considers the "obedience to reality as an exercise of love" (ibid., 41). What this suggests is that attention involves caring about the other's experience, in that a loving gaze upon reality means attending to the feelings and thoughts of the other person from their perspective, i.e., the reality is their reality. Caprioglio Panizza also notes that Murdoch's "[t]ruth is not something passively received, but something we desire, accept, seek, and love. Attention, as a loving gaze, is thus far from passive, but passionate [...] which wants the object to be what it is." (2022, 165). Seeing the other as they see themselves is a characteristic of empathy, or, more precisely, empathetic understanding.

The synthesis of Murdoch's account of attention with Nietzsche's account of value creation therefore gives us a more precise picture of ethical valuation at the

level of the individual, in which a virtue enables the individual to value, and valuation creates a new meaning and, through it, a new disposition. Let us now consider how this process extends from the individual into the social sphere, continuing with Murdoch's example.

6.3.3 From individual disposition to social virtue

Paradoxically, at first sight, it may seem that there is no social good created through M's change in disposition, as no one else witnesses the change, and nothing changes in M's outward behaviour. One might therefore argue that nothing socially relevant has really changed. But this would be to confuse two different stages of value creation. At the level of the individual, valuation reaches its outcome in the formation of a disposition. Whether the value expressed in that disposition is then realised in the relationship between M and D, or more broadly in the social world, is a separate matter.

M's outward (i.e., social) behaviour was appropriate from the beginning, so what changes is the relationship between her behaviour and her inner perspective, bringing the two into greater coherence. The internal struggle she undergoes is therefore not itself the good, but the process through which the new disposition is formed. M continually tests her own perspective against what she takes reality to be, rather than simply performing the right behaviour for instrumental reasons, such as maintaining social harmony.

The notion that disposition is the point at which the individual process of valuation is complete does not mean, however, that an individual's disposition is socially inert. A disposition formed through valuation at the personal level can become a *virtue capacity in the social sphere* because it changes what the individual is capable of bringing into relationships with others. In this sense, an individual disposition becomes the starting point for a further process in the social world, namely a *social virtue* as a capacity to create a social value through our relations with others. The individual disposition formed through valuation is therefore not the end of the ethical process in a broader sense. It is the end of valuation for the individual, but it is also the beginning of what that individual can contribute to the social world.

Let us now illustrate how an individual disposition to see and interpret what matters in a particular way can be brought into relation with others and thereby become a social virtue by examining the history of palliative care. The following account draws on Dame Cicely Saunders's (1918–2005) short description of the history of the emergence and development of the hospice movement (2000), of which she is widely credited as a founding figure. I will then interpret this history from the virtue perspective.

The development of modern palliative care did not emerge because there was a moral demand for it, but as the gradual social expression of a disposition formed through attentive engagement with individual experiences of dying. It began in 1948, when Saunders met a Polish Jewish refugee who was suffering from terminal cancer and whose situation inspired her to rethink care for terminally ill patients. After his death, his £500 inheritance supported her vision of a new kind of hospice. As a doctor, Saunders later developed ideas about pain relief and holistic end-of-life care, while also researching pain and symptom control and raising support for a new institution combining patient care, research, and education. This resulted in the construction of St Christopher's Hospice, which opened in London in 1967 and subsequently became a pioneering model of modern hospice care, later expanding to include outpatient services and home care (Saunders 2000, 7–8).

Saunders's encounter with this patient thus became a starting point for a much broader transformation in how terminally ill patients were understood and cared for, expanding the conception of dying beyond either simply keeping the dying patient's body functioning or leaving things to run their course – that is, beyond a merely technical clinical issue. This encounter generated a change in Saunders's perception of caring for the dying, as the patient's experience made visible something that was not adequately captured by the medicalised way of thinking about death. The problem was not simply that dying patients lacked sufficient pain medication, but that their experiences were much more multifaceted, involving, in addition to physical pain, a wide range of psychosocial dimensions, such as emotional and social distress, spiritual concerns, or fear. The later development of the idea of “total pain” (Saunders 2000, 9) conceptualised this broader understanding, showing that suffering could not be adequately understood by separating physical symptoms from the person's experiences.

The virtue involved here can be understood as honesty in the sense developed above, asking the individual to be open to what is actually encountered rather than forcing the interpretation of the encounter within an already established conceptual frame. In a Nietzschean vein, this requires the effort and courage to move beyond the comfort of an established and stable form of neglect and ignorance (Dennis 2019, 66 fn. 42). It means being willing to face difficult truths, recognise when our understanding is inadequate, and change our views in response to what we encounter. In the history of palliative care, this meant taking the experiences of terminally ill patients seriously, even when those experiences did not fit the dominant medical understanding of treatment.

The crucial point is that this change in Saunders's disposition did not immediately yield a social value. Following the change in her disposition towards dying, the concept of care for the dying expanded beyond curative medicine to

consider dying as a distinctive human condition requiring care. This change in disposition opened up a vision in which a different range of possibilities, such as pain relief, continuity of care, psychological support, family involvement, spiritual concerns, the preservation of dignity, and so on, could now be viewed as parts of the same problem rather than as unrelated concerns (Saunders 2000, 9).

The disposition, which begins as an individual experience, therefore became a capacity for a social virtue. Saunders's changed vision of care for the dying motivated her to initiate broader social change, including developing practices that others could participate in, learn from, and modify, as well as developing methods for gathering more experiences and continually enhancing hospice care (Saunders 2000, 10). In other words, the disposition enabled the changed concept of caring for the dying to acquire a life of its own. What began with one person's willingness to attend honestly to the experience of dying – the disposition of an individual – became a social virtue because it became a shared capacity, unowned by any particular individual.

Saunders's example also shows that social virtues work precisely because they are not tied to one particular principle, however well that principle may have been developed in theory. And they work because they remain open to new experiences and change. The early hospice movement did not simply develop a fixed solution for good end-of-life care and then begin reproducing it; its practices evolved through encounters with patients, research, the sharing of professional experiences, and different cultural contexts (Saunders 2000, 10). Saunders describes the movement as a kind of "kaleidoscope" (ibid., 7), capturing how different experiences and possibilities continually influenced its development. In this sense, the willingness of individuals to reconsider and revise established understandings is what makes the concept viable in the social realm, where it has a life of its own, uncontrolled by any one person, and is therefore self-sustaining. At the same time, it is also self-creative, akin to democracy's self-creativity through social virtue discussed in 6.6.

6.4 CREATIVITY AS A VIRTUE

The meaning of creativity is, of course, multifaceted, as evidenced by the extensive body of work published in the *Creativity Research Journal*, a journal devoted entirely to the topic. Because this study approaches creativity as a virtue, it is useful to emphasise one central aspect of the broader concept, namely that creativity should not be understood solely in instrumental terms, as a means of producing novel outcomes. As Glăveanu and Beghetto argue, it is "about connecting creativity not as much with the outcomes people produce but what they do and how they live their lives; it is ethical, by recognizing the role of our

encounters with the world” (2021, 79). In other words, creativity is grounded in relational engagement with the world, which requires attentiveness to the perpetually unsettled social world.

To see how creative engagement works as a virtue rather than as a talent or an outcome, recall Iris Murdoch’s example of M and D ([1970] 2014, 16–24), discussed in 1.4.2 and 6.3.2. There, attention, in the sense set out in 6.3.2 and exercised together with honesty, brings about a change in disposition. Let us consider what creativity contributes to valuation in this case. M, a mother-in-law, at first sees her daughter-in-law, D, as rude, but gradually comes to see her as spontaneous and refreshingly simple. Nothing external causes the change: M receives no new evidence, D’s behaviour stays the same, and M is polite to D throughout. What changes is M’s vision – the way she sees D. That change comes through “just and loving” attention (*ibid.*, 36), and through M’s effort to overcome her own illusions and see D “as she really is” (*ibid.*).

At first sight, M’s shift from resentment to a positive view would appear to be corrective, so what makes this a creative move? As I argued in 6.3.2, the problem with her initial judgement was not that she failed to notice D’s behaviour, but rather how she understood it. What changes is the meaning M finds in D’s behaviour. And that meaning is not simply sitting there in the facts, waiting to be found. The behaviour alone cannot tell us whether D is being rude. To see that it might mean something else, M needs two things: the honesty to doubt her first judgement and the imagination to think of other ways of interpreting what D does. So, when M comes to see D’s behaviour as ambiguous, her creativity enables her to go beyond what she can immediately observe. If the facts alone settled that D was rude, M’s change of view would be nothing more than the correction of a factual mistake. But nothing about D’s behaviour has changed. What has changed is how she sees it.

Importantly for creativity, attention cannot be required of anyone. M’s internal struggle is prompted by her circumstances, but no external pressure compels her to undertake it. Before reconsidering her interpretation of D, M was polite and friendly because she was conforming to social convention, yet this was a performance. Her behaviour and her beliefs were not aligned, and their alignment was not demanded of her, since convention required only the behaviour. Thus, convention produced compliance, whereas creative revaluation produced a genuine change in value, precisely because it was not witnessed by anyone. Creativity, then, is a virtue-capacity activated by intrinsic reasons, not instrumental gains.

It may be objected that this account permits creativity to justify fantasy and renders attention a mere fabrication. However, Murdoch’s attention actually

resembles realist and agonistic approaches in its commitment to remaining truthful to the actualities of the circumstances, that is, apprehending the world as it is rather than constructing illusions. Truthfulness, understood this way, concerns both matching the phenomenon one encounters and examining one's prior valuation of it. It is not a commitment to apprehending the social world in a form that can be evaluated as true, i.e., objective and universally intelligible, since, as concluded in 6.2.3, evaluating good and bad in this way is analytically unproductive.

Understood in this way, creativity as a virtue is not oriented towards fabricating fantasies that evade reality's constraints; rather, it works with features of the world, recombining and reinterpreting them playfully, i.e., without formal, unsubstantiated constraints in the particular case. It is play in the Nietzschean sense: the creator is an artist who engages with reality not to alter its fundamental structure – an endeavour incompatible with *amor fati* – but to shape its surface, as Miyasaki (2022a, 2) describes; or in the terminology of this study, to shape disposition. The act of play therefore changes not ontology but meaning, through what Nietzsche calls “esteeming” ([1883] 2006, *On a Thousand and One Goals*).

Drawing on Nietzsche, we can identify another feature of creativity, namely its role as the source of *uniqueness*. Recall the description of the Overman in 5.2.2, where the Overman embraces *amor fati* and the will to power, creating themselves as “new, unique, incomparable, who give themselves laws, who create themselves” (Nietzsche [1882] 1974, 266). Yet this uniqueness materialises only when creativity is actively exercised, since a capacity never directed at anything remains dormant and generates no disposition (6.2.1), and this does not happen on its own; it must be cultivated. Nietzsche warns that a person who does not cultivate this capacity settles into mediocrity. His figure for this is the Last Man ([1883] 2006, Prologue §5), who prefers comfort to effort and repetition to risk, and who may well be comfortable precisely because his position in society is a good one.

In turn, comfort of this kind can easily turn into moral domination, as someone content with how things are wants others to conform to the same structure and values. Stability then becomes a demand for certain arrangements, and to maintain it one has to envisage and strive for the settling of a certain order, a world in which sameness is valued. In Murdoch's terms, this amounts to an illusion, a consoling picture whose effect is to remove any reason to look again ([1970] 2014, 17).

Further, Nietzschean uniqueness is intrinsically significant and cannot be subordinated to efficiency, since efficiency, by design, privileges uniformity over originality. The problem with an efficiency- and progress-oriented rationality,

however, is not the use of means to achieve ends as such. A means-end structure does not, by itself, justify domination. The problem arises when progress is treated as an open-ended process, since there is always more progress to be achieved. A goal pursued with perfect efficiency therefore has no internal limit, and its pursuit can extend to the elimination of whatever stands in its way. Obstacles become problems to be removed rather than things to be worked with, producing what Miyasaki calls quantitative domination (2022a, 77), in which one side gains only what the other loses.

By contrast, Miyasaki's (2022a, 77–88) reading of Nietzsche's conception of the will to power describes a qualitative feeling of power (Nietzsche [1888] 1923, 43), rather than a quantitative accumulation of power. The latter is bound to disappoint, not because its aims are unreachable, but because achieving them would bring the process of overcoming, in which the feeling of power consists, to an end (see also 5.2.1). Read in this way, the Nietzschean position warns against a politics oriented towards domination, including politics pursued in the name of a universal good such as justice (Miyasaki 2022b, 97).

On the other hand, accepting struggle and indeterminacy has the opposite effect – if nothing is settled, then oneself and one's world remain open to being made otherwise, and it is friction rather than ease that keeps the will to power in motion. Creativity therefore has two sides, namely refusing to become like others, and attending to what in particular moves oneself, since it is one's own sensibilities and experiences that determine which parts of the world one finds worth engaging with at all.

Creativity, then, is not a matter of moving towards a goal such as justice, no matter how benevolent it is. Rather, it is a matter of imagining what else might be possible, while accepting that imagining carries the risk that nothing comes of it, but this does not make it detached from reality. What one creates with are the obstacles one actually meets, and what matters is the engagement with them, regardless of whether or not they are eventually overcome.

Further, Nietzschean valuation is inherent in creation, as meaning is made through the act of "esteeming" ([1883] 2006, On a Thousand and One Goals). How, then, do we tell whether creativity has been employed? Not by looking at what it produces, but by the valuing itself, which is something felt rather than measured, and so is not something that can be lined up against anyone else's creativity and compared. To clarify, let us compare valuation with normative moral reasoning, which aims to settle what the correct course of action is by using binary evaluation, e.g., good-bad, useful-useless, permissible-impermissible. In contrast, creative valuation is not asking which side of such a pair something

belongs to; rather it is asking whether the thing means anything at all to the person doing the valuing.

A critic might note that this may look like an invitation to disorder, for if there is no fixed aim, no improvement being aimed at, no picture of the good being worked towards, what stops people and societies from simply drifting? While this worry is reasonable, it is also premature, since, as noted earlier, engaging with an obstacle is only constitutive and enjoyable when the obstacle is one a person can actually take on, in thought and in feeling alike (Nietzsche [1888] 1923, 43). And making meaning is itself a way of making order, but in creative engagement the order created is never finished; it is constantly remade as things change and understanding moves on. What makes creativity valuable, then, is not the order it produces but the experience of engaging in the process of trying to produce it, and of being aware of oneself doing so, both rationally and affectively.

Thus far, our discussion has concerned a valuer and her own dispositions. As discussed in 6.3.3, an individual disposition can become a capacity in the social sphere once others can draw on it, and in Murdoch's example, M's revaluation, in this narrower sense, generated such a capacity. But it did not generate a social value, because a disposition becomes a social value only once it is taken up by someone else (6.2.4), and M's change was shared with, and acknowledged by, no one. What happens when creativity-enabled dispositions are taken up by others, and what this requires of political arrangements, is the subject of the next section.

6.5 WHAT IS VIRTUE-ETHICAL POLITICS FOR?

Before sketching a creativity-centred, virtue-ethical approach to theorising about political arrangements in the next section, it is worth setting out the central questions such theorising addresses. This subsection briefly outlines how moralism, realism, and agonism answer the question of 'what is politics for?', and how each conceives political power, so as to provide comparative context before proposing the virtue-ethical view.

6.5.1 Moralism, realism, agonism on politics and power

For Rawls, politics is for securing justice, and stability is what justice produces when citizens share a sense of justice (1996, 103–104, 391). Realists agree on the autonomy of the political sphere (Galston 2010, 390; Rossi 2019, 639) but disagree on what the political sphere contains (Sleat 2018, 2). For Williams, politics is for answering the "first political question", namely the securing of order, protection, safety, trust, and the conditions of cooperation (2005, 3–4). For Geuss, politics is about "agency, power, and interests and the relations among these" (2008, 25),

and the theorist's task is to ask "Who <does> what to whom for whose benefit?" (ibid.). Its purpose, so far as theory is concerned, is diagnostic, namely to see how institutions actually operate and what actually moves people (ibid., 9), by dispelling the illusions that obscure this (ibid., 11). In the agonist camp, for Honig, the purpose of politics is to keep destabilising settled arrangements and to give voice to the *remainders* whom stable arrangements depoliticise ([1993] 2023, 5). For Connolly ([1991] 2002), politics is the arena in which an ethos of agonistic respect and critical responsiveness can be cultivated by agents themselves, so that difference is met with responsiveness rather than viewed as otherness, and it is the conversion of difference into otherness that Connolly's agonism stands against. For Mouffe (2000), politics is for converting enemies into adversaries through institutional design.

However, each of these three strands addresses foundational and persistent diversity differently, approaching politics from a different angle. For Rawls, politics has a clear aim, namely to establish principles of justice that citizens can accept from within their own doctrines. This means that political stability can rest on a genuine agreement, not just a temporary compromise. But for realists, political concerns such as order, legitimacy, and cooperation come first, rather than emerging from a moral theory. Agonists see conflict as a constitutive part of politics rather than as something to be eliminated, instead asking how disagreement can be turned into a form of political contest in which people recognise one another's right to disagree.

Power can be considered central to modern political philosophy, whether theorised explicitly or not. Rawls never conceptualises it directly, but he speaks of political power as coercive, a formulation that treats power as inevitable but largely undifferentiated, that is, as a zero-sum matter, with no attention to its various forms (e.g., power over, power in, power with, power for, power from) or to adjacent concepts such as empowerment. The society Rawls envisages is accordingly one in which power – exercised by citizens as a collective body (1996, 216) – is subordinated to justice (1996, lxii). For realists, politics is essentially about power, as the Geuss quote above in this section shows, and politics can only constrain it. Agonists speak of power in terms of hegemonic order (Mouffe 2013, 14), the suppression or absorption of resistance, but also in terms of disruption (Honig [1993] 2023) and disciplinary power (Connolly [1991] 2002, 21–22).

Thus, for Rawls, power is coercive and politics subordinates it to justice; for realists, power is ubiquitous and self-justifying, and politics is where it must be held accountable; for agonists, power is productive of subjects and political order, and can therefore only be redistributed or disrupted. What unites these otherwise

divergent accounts is that each conceives power in terms of positionality relative to others.

6.5.2 Power and politics: the virtue-ethical view

As discussed in 5.2 on Nietzsche's virtue ethics, upon which this proposition builds, on the personal level, embracing the possibility of engaging with resistance in a way that yields a feeling of power in overcoming obstacles animates human life from within. At the social level, such resistance becomes possible through the diverse perspectives on what living well means, which supply the forum for engagement and create the arena for *adversarial sparring* (4.2.3), aimed not only at understanding the other better, but, through encounter with antagonistic difference, at leading individuals to attend to their own commitments more deeply.

Given the focus on sufficient resistance as the condition for individual thriving in the social world, the virtue-ethical approach accordingly views politics as the arena for creating the conditions of healthy resistance under which engagement with diverse perspectives can be exercised. Such resistance is required to sustain the capacity for virtue, because one's capacity can be exercised only where others retain theirs, and those capacities cannot be suppressed or removed. This rightly raises the question of what happens when unhealthy resistance – one that seeks elimination – is encountered, which I address in section 6.6.

In its focus on resistance, the virtue-ethical view of politics stands close to agonism in form, but differs from it in content: agonists preserve healthy contestation and resist domination or technocratic solutionism as ends in themselves, whereas virtue ethics does so because value creation requires it. The avoidance of domination and solutionism is, on this view, likewise desirable, but only as a derivative good, not as the primary aim.

Thus, the emphasis of virtue ethics falls on conditions, not on a particular design or set of arrangements to guide the adjudication of differing views on political action and social choices (e.g., whether or not 16-year-olds should be eligible to vote), since virtue ethics is deliberately thin in its approach to theorising. In section 6.6, I will sketch a creativity-centred perspective on one aspect of democracy – self-correction – and will reconceptualise it as self-creation, but the uptake of, for example, institutional design is left for future theorising.

Such conditions are, in the Nietzschean vein, what I conceptualise as conditions for empowerment: one's capacity to exercise one's virtues depends on others retaining theirs. By contrast, realists and agonists alike associate power with domination – whether as a property held by someone and exercised over others (e.g., Williams 2005, 5–6), or as relational and productive of subjects and political

order (e.g., Mouffe 2013, 14; Honig [1993] 2023, 3; Connolly [1991] 2002, 21–22). What unites them is not a model of possession but a measure. Power is registered comparatively, as power over another, and it is this that Miyasaki distinguishes as quantitative power, as against the qualitative “feeling of power” (Nietzsche [1888] 1923, 43) in the Overman’s striving for excellence (see 5.2.2).

Let us now revisit a possible objection that this kind of qualitative power and excellence invites criticism of elitism and aristocratic ethics, which I briefly touched upon in 1.3. A critic might argue that the *feeling of power* does not protect against domination over others because a tyrant can easily experience quantitative power in qualitative terms, as without constraints on domination, a power-seeker may be inclined to use whatever means necessary to obtain that feeling, including through overt domination. So, what stops the creative Overman from becoming an anti-democratic elitist, or even a tyrant?

The response to the elitism implicit in the feeling of power is broadly grounded in Miyasaki’s interpretation of Nietzsche, particularly the idea of a “rough equality of resistance” (2022a, 93 fn. 40). This entails that, in principle, the obstacle must be capable of being overcome without being eliminated altogether, as elimination would remove the agent’s possibility of engaging with it. The enjoyment of engaging with resistance therefore emerges when it is strong enough to require the exercise of power, but not so strong that it cannot be overcome.

The question, however, is where this resistance is located. Does it arise in decision-making, deliberation, inclusion, the distribution of resources, or elsewhere? While resistance is something that must be present at every level of political engagement, it is especially crucial at the level of capacities, as there it prompts the agent to initiate reflexivity at the early stage of valuation. If resistance emerges only later, once commitments or dispositions have already formed, it is more likely to produce conflict than contestation. This is because engaging constructively with resistance requires clarifying the reasons for particular commitments and dispositions, and when resistance is raised only after those commitments have formed, their underlying reasons may already be obscured.

Crucially, this resistance is always dynamic, so engaging with it never stops – unlike Rawls’s reflective equilibrium, where adjustment ceases once coherence is reached. To be fair, Rawls allows that equilibrium is unstable and may be upended by further reflection or new cases ([1971] 1999, 18); yet each renewed adjustment moves towards a new equilibrium. Rough equality of resistance, by contrast, is maintained precisely to keep the resolution open.

Recall that virtues are understood as capacities carrying no evaluative content of their own (6.2.1), and thus there is no *higher* kind of virtue or valuation, because

there is no standard against which such kinds could be ranked. Further, when diversity is conceptualised as diversity of reasons (6.3.1) – linked to virtue as capacity – it becomes the underlying assumption of rough equality, since here, too, the reasons lack a standard for ranking.

I call this permanently open form of resistance, which requires parties to remain within the limits of engageable contestation, *constitutive resistance*,¹⁶ to indicate that such resistance creates the possibility of valuation in the first place; without it, there is nothing to value, only something to accept as imposed. This brings us back to the question posed above: what, on this view, prevents the creative Overman from becoming an anti-democratic elitist, or even a tyrant?

Miyasaki's answer is to calibrate resistance against difficulty, so that an obstacle should be neither too easy nor too difficult to overcome (2022a, 94). Virtue-ethical political resistance, however, is not measured against obstacles in general but against other people's capacities, since, on this view, resistance in the social world arises from human relationships (see 6.3.3). Once resistance is understood in this way, it becomes dependent on the continued existence of the other person's capacity to resist. An obstacle created by another person's resistance remains within Miyasaki's 'rough equality' only as long as that person retains the capacity to resist, whereas domination would remove it.

Domination, then, does not simply make the obstacle too easy to overcome; it eliminates the resistance altogether by suppressing the other person's capacity to resist. This suggests that sustained domination may be more difficult to maintain in a political order organised around empowerment than the objection implies. A tyrant with quantitative power does not necessarily experience the qualitative feeling of power in the Nietzschean sense (which is not the same as the lay meaning of the term, see 5.2.1), because such feeling depends on having an adversary who can provide the resistance. This does not mean that formal safeguards against abuse of power are not relevant in real life; rather, it is merely an argument that, on this account, domination undermines the very condition that makes qualitative power possible.

On this view, then, the Nietzsche-inspired virtue-ethical account does not treat power – in the traditional realist terms familiar since Thucydides – as the central feature of political life. Instead, it understands politics in terms of *empowerment*, meaning that politics should create the conditions and capacities that allow individuals to experience a feeling of power. This is not to deny that *power over* is

¹⁶ This response should be read as a defence of qualitative power within my account of virtue ethics, not as a defence of Nietzsche, who indeed reads as an elitist (Owen 2002); my account is Nietzschean, rather than a reading of Nietzsche himself.

part of politics, but rather to say that virtue-ethical politics does not treat controlling power as the first political question.

Thus, the purpose of virtue-ethical politics is to create the conditions for individuals to develop and exercise their capacities, by ensuring the viability of constitutive resistance. Empowerment arises from valuation, and resistance therefore means the sustaining of the conditions of valuation. To be sure, this does not mean the promotion of human development, flourishing, or capability in the sense of Sen (1993) or Nussbaum (2011). Where material conditions enter, they are treated as conditions of valuation rather than as components of a good life.

6.6 VIRTUE-ETHICAL POLITICS

This section sets out what an *ethos* of creative valuation asks of those who theorise democratic politics. An *ethos*, as I use the term, is a normative background orientation, that is, the values and ethical commitments that a theory or framework carries, whether or not it declares them. An unstated *ethos* still shapes how competing positions are judged, while escaping judgement by the same standards. Realist theories have an *ethos* that is anti-utopian and rationalist (e.g., Geuss 2008); agonist theories, by contrast, are marked by an “*ethos* of contingency” (Selg 2012) that seeks to challenge whatever principles of order have come to dominate.

So, insofar as we agree that no social or political theory is entirely value-free, we can say that every theory has an *ethos* that needs to be made explicit, as one cannot construct a normative view that sets certain values as its aim without being clear about the underlying values of that normativity.

My proposal for how an *ethos* of creative valuation can inform normative theorising about politics concerns thinking about democracy as a self-creative enterprise. What follows in this section is based on the argument that an order that secures itself by removing friction becomes fragile, since it eliminates the resistance through which its members work out what their commitments mean. This claim is developed in the following steps: 6.6.1 discusses the conception of politics as constitutive resistance and individual empowerment; 6.6.2 reframes democratic self-correction as self-creation; 6.6.3 grounds this reframing in a conception of democracy as practice; 6.6.4 tests the proposal against the concept of stability; and 6.6.5 justifies the move from rules to considerations and minimal monitoring.

6.6.1 Politics as constitutive resistance and individual empowerment

The process of valuation at the individual level is completed in a disposition, which then underlies our activities in the social world. Individual dispositions shape how we perceive the social world, our place within it, and our relations to others, thereby serving as the basis for social virtue. Much like an individual virtue, a social virtue is also activated by attention, and when valuation produces a social virtue that is practised in the social world alongside other kinds and contents of valuation, this can generate a social value. In section 6.2.4, I discussed this process through the example of truthfulness and trust.

So, what demands do virtues place on citizens? Since virtues are capacities, and the process of valuation requires only that people value, not that they value well, the success of valuation is not measured against any predetermined standard. Moreover, if we agree with Nietzsche that humans are *esteemers* ([1883] 2006, On a Thousand and One Goals), valuation is part of the human condition. Virtue ethics therefore considers all citizens to possess the capacity to value.

By contrast, Rawls's demands for political competence are higher, since they require the moral powers involved in forming a conception of the good and a sense of justice (1996, 103–104), as well as intellectual powers of judgement, thought, and inference and the capacity to recognise the burdens of judgement (1996, 81). Rawls also requires a “duty of civility” (Rawls 1996, 217), understood as the ability to distinguish publicly justifiable claims from private convictions. Realists remain largely silent on this question. Among agonists, Mouffe requires an agonistic disposition to be already formed (Selg 2012, 96–97), Honig assumes an ethos of adversarial sparring (see 4.2.3), and Connolly requires the capacity for self-revision, that is, the ability to “eventually modify what they are, in form or content or both” (1995, 184).

The virtue-ethical account comes closest to Connolly's demand for a willingness to revise one's commitments. However, while Connolly treats revision as an end in itself, this study considers it a source of empowerment that arises through engagement with resistance. The demand placed on citizens is accordingly a thin one, asking only that they not be entirely indifferent to their political surroundings.

Should politics, then, foster the development of virtue capacities? The short answer is ‘no’, since every citizen is taken to possess the capacity to value already. Thus, what politics can foster is conditions for engaging in valuation, but to emphasise, valuation on the virtue-ethical account concerns ethical thinking about relevance rather than judgement on a scale of good and bad, or permissible

and impermissible. And relevance is understood in terms of the Nietzschean esteeming as meaning-making, and this process issues no axiological verdict. The requirement to value is therefore available in principle to anyone capable of empathetic consideration of others, excluding only those in whom other-regarding empathy is wholly absent (see also [Slote 2010, 51 fn. 12](#)).

For others, if we consider valuation of relevance to be initiated by encountering engageable difference, then politics is concerned with sustaining constitutive resistance. Politically created structures are not established to teach anyone how to value; their purpose is to maintain the circumstances under which difference remains available to be valued against. In this context, I have discussed the concept of *rough equality of resistance*, in Miyasaki's ([2022a, 93 fn. 40](#)) terms, denoting conditions in which adversaries are worthy of each other's strength. It is in such conditions that the joy of engagement animates human valuation through encounters with difference, the yet unknown, and anything that appears both capable of and in need of being overcome. I have termed this constitutive resistance (6.5) to emphasise that resistance enables valuation in the first place – in the sense that resistance is what animates attention. Our attention is activated when we encounter a puzzle, ambiguity, or outright conflict with our own beliefs – in short, something that we do not fully understand.

Such resistance emerges naturally from diversity and perpetual disagreement in the social sphere. Like moralism, realism, and agonism, the virtue-ethical approach considers this diversity to be a defining feature of human social life. Viewed through Nietzschean eyes, this resistance is precisely where one can find the will to power and become more than oneself through the qualitative experience of overcoming (these concepts are discussed in 5.2.1). Thus, it is imperative that politics seek not to suppress or control disagreement, no matter how conflictual, but to ensure conditions in which disagreement can provide resistance that is capable of being overcome, and in which all engaged parties can experience it as such – empowered by roughly equal resistance.

The central political concern, therefore, is the individual sense of empowerment, understood in a broadly Nietzschean sense as the drive, impulse, and joy of living. This view of politics does not treat order and stability as the primary goals or goods of social life, since, on the virtue-ethical view, when these become primary, they can constrain the human empowerment that depends on the drive towards creativity. This does not mean, however, that power, stability, or order disappear from the theory altogether. Rather, it emphasises the need to distinguish between power as domination, where power is exercised over others, and empowerment, in which political conditions enable individuals not only to develop and exercise their capacities, but also to sustain the inner drive and motivation for it.

6.6.2 From self-correction to self-creation

Having set out politics as the sustaining of constitutive resistance, we can now ask what follows for self-correction, as a core feature of democracy, and how it can be reframed in terms of virtue-ethical self-creativity. In particular, I ask whether democracy can be understood not primarily as a mechanism for determining the democratically correct collective outcome, but as a political arrangement that aims at providing conditions for continuous contestation and revision-friendly valuation. Importantly, such a creatively self-corrective – that is, what I will call self-creative – political order should not mean correcting political decisions against a predetermined standard of the good, but rather allowing the good to emerge from within the process of valuation.

In broad terms, self-correction is the capacity to identify and rectify errors by relying on internal resources rather than external impulses. A self-correcting society has mechanisms for recognising symptoms of failure, diagnosing their sources, and implementing remedies. Self-correction is one of democracy's defining features, and its mechanisms are commonly conceived in institutional terms, structured through procedures designed to ensure continuous renewal, such as regular elections and accountability frameworks (Przeworski et al. 1999).

Yet these institutional mechanisms of self-correction risk becoming rigid if their purpose is conceived as correction in the service of stability, where the emphasis falls on the *correct* rather than the *self*. Correction on its own implies the restoration of a predefined standard, where there is a wrong against which something is corrected and an understanding of what the correct situation should be. By contrast, 'self' reflects agency and its active disposition towards change.

Such institutional self-correction tends towards rigidity where its corrective function carries excessive weight, since the emphasis then falls on strategic and calculative reasoning alone, that is, on instrumental rationality. Oakeshott's (1962) critique of such rationalism provides a helpful explication. He argues that it prioritises the universal over the local and principle over practice, culminating in what he terms the sovereignty of technique, grounded in "the pre-occupation of the Rationalist with certainty" (ibid., 11), which emphasises technical knowledge, set out in rules and propositions, over practical knowledge, which is acquired through experience and cannot be rigidly formulated (ibid., 11–13). Where judgement is assumed to be exhaustible in data, observation, and calculation, experience is excluded, and correction becomes a matter of adjusting rules or procedures rather than conceptions, interpretations, meanings, or relations. Accordingly, closure of preset aims and predictability become the guiding light, and disruption becomes something to be managed rather than engaged with.

Consequently, rationalist views have little to do with the ‘self’ in self-correction, including foundational value commitments, and much to do with correction – that is, the adjustment of social and political roles and arrangements. Because they rely on existing standards of justification, they lack the resources to move beyond prevailing value frameworks that are affirmed via public reason and legitimacy, and thus remain confined to institutional correction. This view applies to approaches that require a standard of correctness that is independent of, and prior to, the democratic process. For instance, Landemore (2013) argues that cognitively diverse groups identify better and worse responses more reliably than non-democratic procedures, whether coming from a council of experts or a benevolent dictator.

Agonism, by contrast, is primarily oriented towards the ‘self’. Agonist self-correction is inherently paradoxical, since it rejects stability as an end, yet requires reflexive self-regulation to prevent instability from becoming disabling (Honig [1993] 2023, 204). However, agonistic self-correction is largely directed towards sustaining agonism itself. This is especially evident in Mouffe’s work, where agonism retains structuralist and stabilising features, such as group-based identities (4.4.1). Honig (4.2) and Connolly (4.3) are more disruptive, embracing pluralism and open-endedness without substantive frames, and it is precisely this agonistic paradox – the tension in which agents must engage resistance while seeking a provisional balance that remains open to contingency – that lies at the heart of self-correction as such.

This prompts a question: if agonism already emphasises disruption, open-endedness, and the exercise of agency, why do we need a virtue-ethical alternative?

Let us first consider where divergence in value commitments actually occurs. On a decision-centred view, disagreement emerges when agents reach different conclusions. Yet two agents may reach the same conclusion for very different reasons, while two agents who share the same foundational commitments may reach opposing conclusions because they interpret the situation differently. The source of the divergence therefore lies prior to decision – in how agents conceptualise the problem. By contrast, on the virtue-ethical account, such conceptualisation is itself a form of meaning-making through valuation.

This is why I argue that value commitments diverge not primarily at the level of verdicts, but at the level of how a situation is problematised. Elements of this epistemic dimension have been recognised by realists (Williams 1985) and, more extensively, by agonists (Connolly [1991] 2002; Honig [1993] 2023), and although agonists draw on Nietzsche to frame struggle as central to political action, they overlook his emphasis on esteem and creativity as forces that animate individuals

and motivate political engagement, as resistance can itself be experienced as empowering. By contrast, a virtue-based approach enables the creation of what does not yet have a standard. Understanding how values get their meaning requires understanding the foundational value concepts through which we interpret and engage with the world. Agonists show that contestation animates political life; the virtue-ethical account explains why it is so, locating the joy of contestation in the exercise of a capacity against resistance.

This hopefully explains what the virtue-ethical account seeks in self-correction and why I propose self-creation as a better term. The emphasis on creation marks the process of *correction* not as a return to a previous or approved state (as per a certain standard), but as the creation of a new meaning within oneself. Because what is created is genuinely new rather than simply restored, change cannot come from correcting an earlier judgement. Instead, it must happen within the process of forming judgements itself, since this is where something genuinely new can emerge rather than simply reproduce an existing standard. In other words, self-creation is about changing how we form value judgements, not just correcting the mistakes those judgements produce.

6.6.3 Democracy as practice

Having outlined the rationale for moving from self-correction to self-creation, we can now turn to the argument in greater detail, beginning with two thinkers who understand democracy as a living practice that is constantly re-created by its members: Dewey and Arendt. Let us begin with Dewey's characterisation of democracy in his 1939 essay "Creative Democracy – The Task Before Us" before unpacking it.

"Democracy is a way of personal life controlled not merely by faith in human nature in general but by faith in the capacity of human beings for intelligent judgment and action if proper conditions are furnished. I have been accused more than once and from opposed quarters of an undue, a utopian, faith in the possibilities of intelligence and in education as a correlate of intelligence. At all events, I did not invent this faith. I acquired it from my surroundings as far as those surroundings were animated by the democratic spirit. For what is the faith of democracy in the role of consultation, of conference, of persuasion, of discussion, in formation of public opinion, which in the long run is self-corrective, except faith in the capacity of the intelligence of the common man to respond with commonsense to the free play of facts and ideas which are secured by effective guarantees of free inquiry, free assembly and free communication?" (Dewey [1939] 2021, 63)

Although written in 1939, Dewey's pragmatist account remains strikingly contemporary in treating politics as an ongoing practice grounded in a commitment to the commonsense character of human beings. Democracy, he writes, is "a personal, an individual, way of life [and] a way of life controlled by a working faith in the possibilities of human nature" ([1939] 2021, 62). Dewey's democracy draws on the practical intelligence that leads people to communicate with one another in arriving at shared judgement and action. His account likewise emphasises the "inventive effort and creative activity" (ibid., 61) that become vital in times of crisis and at societal turning points.

Notice what Dewey describes as self-corrective: public opinion, which corrects itself "in the long run" ([1939] 2021, 63). This makes for a narrower claim than the broader idea that democracy as a whole is self-corrective. Dewey does not focus here on democratic institutions (e.g., elections) as mechanisms of correction. Instead, he focuses on public opinion and, therefore, on people themselves, shifting attention from democracy as a system of rules and institutions to democracy as an ongoing activity that people sustain through their own participation.

Dewey grounds democracy in both a human capacity and the conditions under which it can develop: human beings are capable of "intelligent judgment and action if proper conditions are furnished" ([1939] 2021, 63). Those conditions are free inquiry, free assembly and free communication (ibid.). In my reading, there is no independent standard against which such judgement and action are measured. They emerge instead from the practice of democracy itself, through "consultation, conference, persuasion, and discussion" (ibid.), and under these conditions what results is the best that can be expected in the circumstances. Correction, accordingly, is gradual. It unfolds over time, and so the democratic process will inevitably make mistakes in public. Yet these are not errors to be put right from outside by some higher standard; they are part of the process by which democracy corrects itself.

For Dewey, democracy is "a way of life" rather than a "political mechanism that will work as long as citizens were reasonably faithful in performing political duties" ([1939] 2021, 62), i.e., not something that we sustain mechanically by holding elections or following principles that ensure certain rights. It is an internally driven attitude that is inside people, not structures. He says,

"democracy is a personal way of individual life; that it signifies the possession and continual use of certain attitudes, forming personal character and determining desire and purpose in all the relations of life. Instead of thinking of our own dispositions and habits as accommodated to certain institutions we have to learn to think of the latter as

expressions, projections and extensions of habitually dominant personal attitudes.” ([1939] 2021, 62)

Considering that democracy is sustained by people, rather than structures, we can see why holding a fixed ideal of a political order or substantive aims fails to be substantively self-corrective. While elections and other mechanisms of democracy ensure change, such change risks remaining formal and performative, an illusion of change in substantive terms.

To illustrate, suppose a political order assumes a fixed ideal, such as justice, consensus, or a well-ordered society. It is safe to say that, because of the complexity and particularity of circumstances, that ideal cannot be fully achieved. But if democracy is considered primarily as a mechanism of arrangements, set as a goal to achieve, then when the expected result repeatedly fails to materialise, the most obvious explanation is that the institutions and the structure are not working properly. The substantive content itself is left unquestioned, and the response is to adjust the institutions incrementally, in the hope that repairing them will eventually produce the desired outcome. But this expectation is repeatedly disappointed because the problem is not really institutional in the first place, but lies in the illusion of a democratic ideal. On this reading, dissatisfaction with democratic practice is produced by the very ideal that was supposed to justify it.

This echoes Oakeshott’s critique of rationalism discussed in 6.6.2, which argues that designing democratic institutions in advance substitutes technical knowledge for the practical knowledge on which democratic participation actually depends, i.e., for an awareness of one’s own circumstances that comes from experiencing them and cannot be fully captured in rules.

Taken together, Dewey’s and Oakeshott’s arguments open the way for the claim that the institutional form of democracy is best left open. In a Deweyan spirit, people need the space and freedom to take part in democracy in their own ways, because democracy is something people practise in their own circumstances. The theorist’s task, then, is not to define the ideal democratic structure, but to think about the conditions that foster people’s participation in democratic life.

Those conditions involve creating enough intellectual and affective space for people to express themselves. This is the democratic ideal properly understood, as an ethical matter concerned with how people relate to one another. In Dewey’s words, democracy is a “moral ideal” ([1939] 2021, 64) of respect and consideration for one another’s divergent commitments:

“A genuinely democratic faith in peace is faith in the possibility of conducting disputes, controversies and conflicts as cooperative

undertakings in which both parties learn by giving the other a chance to express itself, instead of having one party conquer by forceful suppression of the other – a suppression which is none the less one of violence when it takes place by psychological means of ridicule, abuse, intimidation, instead of by overt imprisonment or in concentration camps. To cooperate by giving differences a chance to show themselves because of the belief that the expression of difference is not only a right of the other persons but is a means of enriching one's own life-experience, is inherent in the democratic personal way of life." ([1939] 2021, 64)

Dewey here echoes what I discussed in relation to Honig (4.2.3) and to virtue-ethical politics, namely that we engage with views that differ from our own, and treat them as resistance, not merely because others have a right to hold and voice them, but because the encounter enriches our own experience, letting us see more clearly where our commitments came from and take a reflexive view of them.

Let us consider now what a focus on creativity adds to Dewey's people-centred account. If we understand democracy not as a structure of governance but as an ethos – as a way of thinking and acting that people develop through engagement with those unlike themselves while still seeking to cooperate – then novelty stops being an accidental by-product of democratic life. Arendt's idea that politics is grounded in natality is a good example of how democracy, understood as an ethos, is an ethos of novelty, that is, the capacity for new beginnings that each of us has simply by virtue of having been born. Action, Arendt writes, is "ontologically rooted [in] the fact of natality" ([1958] 1998, 247), since "the new beginning inherent in birth can make itself felt in the world only because the newcomer possesses the capacity of beginning something anew, that is, of acting" (ibid., 9).

Here, Arendt draws our attention to the fact that what creates novelty should not be understood in some socially aggregated or absolute terms, but rather that each new human being, by virtue of being, has the capacity to renew the social world. She continues with the thought that, rather than some fixed moral aim, it is precisely the capacity to renew the social world that is the core characteristic of political action. As she puts it: "since action is the political activity par excellence, natality, and not mortality, may be the central category of political, as distinguished from metaphysical, thought" (ibid.).

What makes Arendt's renewal possible is plurality, understood as "living as a distinct and unique being among equals" ([1958] 1998, 178). And living here is not meant in passive terms, but as active participation among others. Because we act among others, our own commitments are transformed, so that the outcomes of

our actions cannot be predicted, processes cannot be reversed, and there is no single author or authority over the process.

Novelty matters because it gives us fresh material to engage with, and thus an order that suppresses it narrows what we are able to value. By contrast, openness to the unknown, including its capacity to disrupt and to produce error, is what keeps democracy resilient. However, Arendt also notes that an overly unpredictable society – and any truly plural society is unpredictable – risks producing frustration that makes us want “to replace acting with making” ([1958] 1998, 220), to “do away with the plurality” (ibid.), and to install a “monarchy, or one-man-rule” (ibid., 221) in order to escape the perplexities of the social world.

Such one-man rule can also manifest without one particular individual or within one particular political order. It is not merely about being well-ordered, but about being one-ordered, where structures and rules function as one coherent system with the aim of maintaining stability and coherence. Here, the rules and structures become the priority and a tool for taming the diversity-driven perplexities of social life. When democracy comes to rest mainly on formal structures that people follow, that substitution is exactly what has happened, and the Nietzschean will to power is tamed – the drive to put oneself forward and to feel empowered through engagement with those who see things differently.

Where agents meet one another unconstrained in resistance, however, the danger that someone will turn a temporary advantage into lasting domination is reduced, because domination destroys the resistance on which the feeling of power depends. In this sense, constitutive and roughly equal resistance does not rule out domination per se; it merely shows that domination causes harm and suffering not only to the oppressed but also to the oppressor, because domination ultimately destroys the source of the feeling of thriving.

The task of virtue-ethical politics is therefore to avoid both suppressing people’s capacity to challenge the shared way of life and relying on an extensive system of safeguards. And since regulation alone cannot achieve this, it must create conditions in which people come to care about society as a shared form of life, rather than trying to secure stability or rights primarily through rules. An ethos of self-creative democracy accordingly focuses on the self of its members, in the sense that the ‘self’ of democracy is the ‘self’ of each of its people. Dewey’s account of public opinion and Arendt’s of natality each supply part of that thought, though neither draws from it the kind of rejection of self-correction that I do.

6.6.4 Stability: weak and strong

Just because an order is created by its members rather than guaranteed by its structures does not mean that it must fall apart. Realists reach for stability

through legitimacy; Rawls's well-ordered society rests on citizens continuing to affirm a consensual conception of justice; agonists' stability relies on the health of adversarial relations. While their vocabularies differ, all are seeking some sort of stable and ordered balance. So, what kind of stability, if any, does the ethos of self-creative democracy envisage?

Conceptually, we can distinguish between different dimensions of stability, such as source (who or what ensures it), scope (to whom or what it is applied), and content (what is required substantively for an order to be called stable). Since I cannot discuss them all here, let us focus on the first, and in particular on stability with a source in political agents, since at the centre of virtue-ethical politics is the individual.

Vallier (2017) considers a liberal well-ordered society's stability to rely on three features of agent-based stability: durability, balance, and immunity. Stability is ensured only if they work in coherence and are present for the right, that is, moral, reasons (ibid., 233, 258). Durability is achieved via social trust and cooperative behaviour (ibid., 245); balance denotes low short-run variability of trust (ibid.); and immunity is about resisting destabilisation caused by non-compliant agents (ibid., 249).

Vallier's model is intended to measure different kinds of stability, and my critique targets not his method (as it simply serves a particular purpose), but rather the two kinds of agents he presupposes. Reasonable agents comply with the cooperative environment; that is, if many around them are cooperating, they too cooperate. Merely rational agents, by contrast, simply maximise their utility (2017, 249) and defect from cooperation when it does not serve their interests, thereby making the whole system less cooperative (ibid., 250). At the same time, the system is said to be stable for the agents' morally grounded *right reasons*, while those agents are considered fully calculative and their relationships within society are understood from the point of view of the prisoner's dilemma (ibid., 249).

From the virtue-ethical point of view, Vallier's agents are not autonomous actors but merely part of a cooperative dilemma in which the kinds of actors and choices are limited, and their decision to cooperate is based on pure calculation, without considering the underlying reasons of others. In this model, reasonable agents count how many people already cooperate and behave accordingly (2017, 237), guided by an assumption of risk aversion (ibid., 245) and an overall strategic disposition towards the social sphere. At the same time, the model presumes that reasonable agents choose for moral reasons, which makes morality a means to strategic goals, and therefore trust in this case is also considered in terms of strategic rather than sincere or genuine trust. Further, Vallier implies that

perfect immunity, that is, a society with no unreasonable agents, is desirable, considering such agents akin to Rawls's unreasonable ones who are excluded from the consensus-seeking stability (ibid., 254).

Vallier's purely strategic agents are also something that the virtue-ethical view resists, since cooperation cannot be grounded purely in calculative conformity, but the *right reasons* must be sincerely felt and experienced empathetically towards others. So, while Vallier concedes that his model is simplifying – for example, agents are cognitively limited (ibid., 258) – the model is missing the kind of agent who is creative and considers the situation not only in strategic terms, but as an opportunity to engage with different points of view and to think about their own commitments in the broader context of themselves and others.

The creative agent therefore has a more nuanced take on the situation than simply copying others' choices or removing oneself from it without having any particular fixed goals in mind. Such an agent tries to understand what moves others and to engage their own reasons with those of others. The strategic reading of interests, by contrast, is an artefact of the model itself, since agents appear as strategic players because the system casts them as players in a Nash equilibrium (Vallier 2017, 236). The virtue-ethical view considers this a fictional construct, because it abstracts away the messiness and unpredictability that make human life human.

Considered in terms of durability, balance, and immunity, creative agents would therefore attend to the meaning of people's commitments, that is, to how they came to value certain aspects of cooperation.

As for durability, which is grounded in social trust, Vallier measures it in terms of how much people cooperate, inferring that cooperation brings about trust. From a virtue-ethical view, we ask about their reasons for cooperation, without automatically inferring that it yields trust. Trust can indeed be considered a critical component of endurance, but for people to trust others, they have to trust their reasons. Only then will others' actions be intrinsically predictable and thus genuinely trustworthy. In strategic games, this would amount to trusting the probability of prediction, rather than trusting other people.

In terms of balance, on a virtue-ethical account, enough friction is needed to prompt people to work their differences out. Balance is therefore not Vallier's perfectly flat line, which is not only stagnant and rigid but, crucially, fragile and prone to breaking upon even a weak disturbance. Instead, it is a virtue-ethical form of resilience.

In terms of immunity, Vallier's ideal endpoint is a place where nobody remains who disagrees. On the virtue-ethical account, this kind of society is not a healthy

one, as it has effectively eliminated the hardest resistance, thereby constraining the space for engaging in valuation through encountering obstacles. On the virtue-ethical view, this would therefore not be a triumph but a defeat, since only the Last Man of Nietzsche ([1883] 2006, Prologue §5) would survive, no longer having reasons to figure anything out in terms of social relations. Virtue-ethical immunity would therefore not be free of radically different thinkers; on the contrary, in a way, they would be welcomed, much like a vaccine that activates the immune system of society.

In sum, the virtue-ethical view of stability is not a stabilising one, but rather one that seeks stability through resilience and by opening the possibility of friction. On this ethical view, seeking stability for the sake of stability actually makes it weaker because it tries to eliminate anything that challenges it, thereby making it vulnerable to sudden disruptions. In contrast, embracing seemingly harmful (but not destructively so) agents and ideas helps make stability stronger by constantly pushing it towards improvement.

6.6.5 From supervision and rules to considerations

Since self-creative democracy is not designed to rectify errors, the question arises how it can avoid harmful domination, including abuses of power. More broadly, it raises the question of what such a democracy would require in terms of regulation, since any social order needs arrangements that make collective life navigable. My concern here, however, is specifically with enforcement. My argument is that if democracy is understood as securing the conditions for valuation, rules should be broad and enforcement minimal. This leaves room for creative and contestable interpretation, rather than replacing it with enforcement. Excessive enforcement not only weakens the drive to contest but also reduces the incentive to take responsibility, and can consequently diminish trust in the social world. Yet trust, rather than rights or legitimacy, is precisely what virtue-ethical politics takes to be at the core of democracy.

I therefore propose a distinction between two kinds of democratic regulation: supervisory and coordinative. Coordinative regulation helps people navigate social life, simplifies routine tasks, and streamlines processes. When coordination moves beyond these practical aims and begins to exercise normative power, it becomes supervisory.

The problem with supervisory regulation in a democracy is that it risks undermining trust. Michael Power's (1999) study of what he calls the "audit society" describes the "audit explosion", denoting the expansion of formal verification across different domains (e.g., environmental audits, management audits, forensic audits, data audits, intellectual property audits) into areas that

had previously relied on professional judgement, such as universities and schools (ibid., 3), observing that this resulted in increased performativity. First, practices were progressively adapted to meet auditing standards (ibid., 47), leading to what, in Oakeshott's terms, amounts to the substitution of technical for practical knowledge (6.6.3). Further, audit removed the conditions necessary for fostering trust, since when someone's competence is constantly verified, trust is no longer necessary; audits, on Power's account, "create the distrust they presuppose" (ibid., 135–136). Consequently, audit becomes what Power calls a ritual of verification (ibid., 123), in that it is no longer so much about checking whether an activity is performed well and with care as about assuring the public that the activity has been checked, so that eventually "audit concerns itself with auditable form rather than substance" (ibid., 96).

Onora O'Neill (2003) argues that the accountability revolution, introduced to restore public trust, has instead weakened it. As professional and institutional performance has become more closely monitored, mistrust has increased (ibid., 52). For O'Neill, trust cannot be created through disclosure alone, since it requires people to exercise judgement about whom and what to trust, and transparency may even undermine this process by producing "a flood of unsorted and uncheckable information" (ibid., 79), while, at the same time, professionals spend more time producing records and meeting targets and less time exercising the judgement and engaging with people that their work requires (ibid., 50–51). The result is a self-reinforcing system in which failures of accountability are answered not by reconsidering the system, but by demanding "more of the same" (ibid., 51).

Thus, practices such as monitoring and accountability induce performativity, thereby reducing the incentive to exercise virtues such as sincerity, accuracy, and honesty. As a result, the disposition towards truthfulness may remain weak or fail to develop at all, and citizens may lose the basis for trusting one another. Surveillance can produce correct behaviour that is outwardly indistinguishable from the behaviour of a truthful person. But compliance cannot produce trust because, on a virtue-ethical account, trust emerges from valuation and internal motivation. A disposition ceases to function as a virtue, however admirable or well exercised, when it no longer arises from the agent's own valuation but instead becomes a form of performance (see also 6.3.2). Thus, when truthfulness is demanded by an external authority, it becomes mere compliance because the actor has not engaged in the process of valuation.

So, the case against supervisory regulation does not rest on the infringement of liberty, nor on arbitrary interference, but on the claim that monitoring undermines the conditions under which valuation occurs, and therefore it forecloses the creation of the very goods, including trust, that a supervisory mechanism is called upon to secure.

The same holds for regulation by rule. Limited enforcement of socio-political conceptions means that the content of conceptions such as justice is left undetermined, while the concept itself remains, retaining only the *sense* of justice in the Rawlsian vein (see 2.3.1). Since self-creative democracy requires room for agents to value, its rules and regulations cannot settle questions of valuation prior to agents' own valuation. Thus, democratic political arrangements are better framed in terms of considerations rather than rules. A rule is understood as giving an answer in advance about what one must do, whereas a consideration tells us what to pay attention to while leaving the answer open. For instance, a speed limit is a rule, while reasonable care is a consideration.

The problem with rules is that they are intended to provide an answer to a question prior to the agent's judgement. Although criticising a rule can involve making judgements about its value, such criticism is still largely constrained by what the rule has already determined. In other words, the rule sets the terms of the debate by obscuring the valuation that its application in fact requires. As Wittgenstein points out, adding further rules to clarify how a rule should be applied does not resolve the problem, since "every interpretation hangs in the air together with what it interprets" (2010, §198). What actually gives a rule its meaning is the practice within which it is used (*ibid.*, §202). Thus, a rule without considerations can hide the question that originally gave rise to it. By providing an answer in advance, it constrains how agents can understand and respond to a situation, leaving less room for them to exercise their own judgement.

The problem becomes particularly clear when agents are uncomfortable with the rules governing the social order and those rules are enforced through supervision. When they seek to understand the source of their discomfort, they are likely to focus on the rules themselves and conclude that the rules are wrong. When people feel that they lack the power to change established rules, resentment can grow, and this may take the form of iconoclasm, directed at the visible instruments and symbols of an order rather than at the underlying problem, which is often much harder to see.

By contrast, when social arrangements are discussed primarily in terms of considerations rather than fixed rules, agents have greater scope to exercise their capacity for valuation, since they can ask why a particular rule exists and what meaning or purpose it is meant to serve, directing their attention towards the underlying problem that the rule was intended to address. This follows from my earlier argument (6.6.2) that the source of differences in value commitments is not at the level of moral decisions, but at the level of how problems are conceptualised. Thus, when people disagree about a rule, for example, a vaccination policy, they may reject it for very different reasons. By contrast, when the idea of vaccination is understood in terms of considerations, people are

prompted to attend to the relevant factors and work out the answer for themselves, which gives them a reason to engage in valuation. And when this process of valuation enters the social sphere, it becomes a resistance-driven polylogue.

The same line of thought applies to the notion that virtues should not be viewed as principles (6.2.1). On the virtue-ethical account proposed in this study, honesty, for example, is not the rule “one must always tell the truth”; it is a capacity that, once activated by attention to the situation, its circumstances, and the valuer, gives rise to a valuation from which a disposition towards truthfulness can be formed. Likewise, a consideration orients without determining the outcome. By contrast, a rule works like a principle by providing an answer in advance, leaving the agent with the option of either accepting or rejecting that answer and thereby directing attention away from the underlying question that the rule was originally intended to address.

Considerations are therefore not primarily about choosing between competing concerns. If two considerations pull in different directions, e.g., reasonable care and individual freedom, in virtue-ethical terms, there is no rule that determines which should take priority. Instead, conflicting considerations create the kind of resistance that requires people to engage in valuation, and resolving these conflicts in advance would therefore remove the space for valuation. Consequently, a self-creative democracy must regulate in a way that leaves its citizens something to value. In other words, it must leave open the questions that agents must consider, rather than determining their answers in advance.

Now, if rules are replaced by considerations, the meaning of violation likewise changes. While a rule can be breached because it establishes in advance what counts as non-compliance, a consideration is relevant to the circumstances and to the valuation of the agents involved. Thus, in the case of disagreement about which consideration is relevant, the appropriate response is to inquire into the reasons behind the consideration. In virtue-ethical terms, this means asking what virtues (if any) are involved in creating a disposition to respond to the consideration in a particular way. The parties must set out what they take the consideration to entail and why, and in doing so they explicate and reconsider the values at issue – not only those of the other party, but also their own. Such disagreement is constitutive rather than merely obstructive, in the sense that it is sustained through roughly equal contestation (see 6.6.1), which is what constitutes valuation within an order oriented towards the empowerment of people. Contestation thus functions as part of the process through which the meaning of a consideration is worked out, rather than as an obstacle to regulation.

It is also worth distinguishing a consideration from an unenforced rule. In the latter case, even if a rule is not enforced, it still commands action; it simply does not punish those who fail to follow it. Thus, one knows what action the rule requires, but there is ambiguity about the consequences of non-action. The main point here is therefore that a rule does not call for valuation, since the end is already established, regardless of how strictly the rule is enforced.

Consequently, enforcement should focus only on concrete and serious harms. It may be objected that serious harm is exactly the kind of predetermined substantive good that this account has rejected, but the difference is that a conception of justice, or of a well-ordered society, tells us what valuation should conclude, while the avoidance of serious harm specifies the conditions required for valuation to take place in the first place. So, harm is excluded for the same reason that domination was ruled out in 6.5, as both eliminate the resistance on which the *feeling of power* depends, and if an agent is harmed, they are in no position to value anything. Enforcement aimed at avoiding serious harm therefore protects the conditions of valuation without settling its content.

How would such a consideration-based ethos work in practice? Consider the debate over banning the veil in public spaces. When framed as a rule, the debate becomes a binary choice between permission and prohibition. The veil is treated as the object of regulation, and the broader question becomes whether autonomy in clothing should be subject to public regulation. But this discussion often starts from a meaning that is treated as already fixed, in which the veil is understood as a marker of oppression, as it is typically portrayed in Western societies (Abbas 2023). The veil, however, can have many meanings at once, e.g., various religious, political, and familial ones. There is no single and fixed meaning, and Muslim women themselves differ in what they believe veiling expresses, as also shown by the different ways in which it is worn. Thus, in practice, people attribute different values and meanings to the veil (Scott 2010).

A rule-oriented approach focuses attention on the question of whether the veil should be banned or allowed and applies a meaning based on a broadly shared conception. When the veil is treated simply as an object, however, the agency of the person wearing it is easily left out of the debate. We therefore miss an opportunity to discuss the reasons for veiling and to open that discussion to everyone involved. A consideration-based approach starts with the practice itself and the reasons people have for it, not with a rule decided in advance. The aim is not simply to explain why people wear the veil, but to create space for everyone involved, including the women who wear it, to think about those reasons. That way, they can reflect on their own commitments without anyone having already decided whether the practice is acceptable. As a result, the wearer's own judgement about the veil can be seen as a choice about what they value, not simply

a habit or something imposed on them. Or, if such an imposition exists, this is also revealed through the process of valuation.

In broad terms, an ethos does not require any particular institutional form; it can be expressed through different kinds of institutions. Ostrom's (2000) study of successful long-term governance shows how local institutions can give people opportunities to work out what rules mean in practice, and can be more successful than systems governed from outside (*ibid.*, 33). Three of Ostrom's design principles (*ibid.*, 41–42) are especially relevant here: rules should reflect local conditions and people who live under them have to be able to change them; sanctions should be proportionate to the seriousness and circumstances of the violation; and institutions should provide accessible ways to resolve disputes, since unresolved disagreements about how rules apply can weaken people's willingness to comply.

Such an approach does require investment in public awareness and in fostering citizens' capacity to engage with the considerations at issue. But this is precisely the point of preferring considerations over rules, since rules also require effort in terms of enforcement, but their enforcement can come at a greater cost to the development of people's capacity to engage in valuation, as Power's (1999) account of the audit society and O'Neill's (2003) account of the accountability revolution both show.

This explains how a self-creative democracy can be resilient, for while an order based on rules may appear stable since its requirements are clear and enforceable, this stability can be fragile if people follow the rules without considering why these rules matter, leaving the source of their substantive commitments dependent on an external authority. But when that authority is no longer there or the rules no longer fit the situation, people may lack the capacity to respond adequately, because, as noted above, the first instinct may be to target the rule rather than address the underlying problem. By contrast, an order based on considerations is less rigid because people are driven to continually work out what their commitments mean in particular circumstances, and because they cannot simply inherit the meaning of those commitments, they become active participants in creating it.

Consequently, the apparent instability of such an order can actually make it stronger. Thus, a self-creative democracy cannot eliminate uncertainty without also narrowing the space in which valuation takes place, and its resilience lies in the continuing capacity of its members not only to respond when circumstances change, but also to remain proactively engaged, in the Nietzschean sense of the will to power, with changing circumstances. In short, political arrangements

should be framed in terms of considerations rather than rules, so as to require and reveal more valuation rather than less.

6.7 CONCLUDING REMARKS AND POSSIBLE OBJECTIONS

6.7.1 Core tenets of the ethos of valuation

This chapter proposed a virtue-ethical ethos of theorising about politics. I will now briefly summarise its core ideas and conceptions to remind us of its distinctive features before addressing some possible objections that can be raised by its critics.

My proposal shifts how we think about values from prioritising value to prioritising the process of valuation. On most accounts, virtues are understood as dispositions, i.e., as stable, cultivated ways of responding to situations, whether understood as means to the good or as partly constitutive of it. My proposal separates the two in order to specify the process of valuation.

On my account, the meaning of the good is constituted through the exercise of virtue in valuation, and virtue is understood as a capacity of character (6.2.1), that is, a narrow trait that enables ethical valuation without carrying any evaluative content of its own. For example, honesty and sincerity are capacities in this sense. Virtues remain dormant until attention directs them at something upon encountering a situation that prompts consideration. Attention involves care and consideration towards the situation, its circumstances, and others in it, but also care about one's own commitments and a readiness to face reality honestly, even when (and especially if) doing so requires understanding one's own commitments. The drive that animates such attention is the Nietzschean will to power in its qualitative sense, namely the impulse to engage with what resists. This is why attention is prompted by situations in which others express commitments that differ from our own, or that are otherwise puzzling. And since society is persistently diverse, such encounters abound.

The set of virtues that engages with a situation perceived as requiring some kind of valuation because it is (at least to some extent) unknown or simply feels alien, through the process of valuation (6.3.2), forms a disposition (6.2.2), which is a way of seeing rather than a rule for acting. The process of valuation creates meaning for the situation; i.e., it is not a verdict (good/bad), but uses a narrative mode of thought, one's own story about the situation, which then makes up the substance of the disposition. A disposition towards a consideration in the social sphere, i.e., upon encountering other perspectives, becomes a capacity of social virtue (6.3.3), which ultimately, through relations with others, creates meaning for social value

(6.2.4). To reiterate, valuation is understood here in terms of relevance and meaning, rather than as a final judgement; it is first and foremost about what appears relevant to the case in relation to one's commitments and sources of motivation. In short, valuation negotiates the external world with the agent's inner experience, and since we are here discussing ethical valuation, it is framed by the question of 'how we are towards each other?' (1.4.2), both in an affective and cognitive sense. Thus, valuation is a complex mix of relationality, reflexivity, affect, and internal motivation, and the thought process runs in a narrative form.

To be sure, virtues as raw epistemic capacities have no fixed content on this account. The terms used here therefore describe positions within a process rather than fixed meanings. Moreover, valuation is not a linear process. Presenting it as a sequence from attention through virtue to disposition and social virtue is useful in analytical terms, but in practice these stages influence one another, e.g., an existing disposition shapes what we notice. Thus, my account is concerned with the process of ethical valuation rather than with defining what any particular virtue must mean.

This process can be seen by looking at how a personal disposition becomes socially meaningful. For example, a disposition towards truthfulness can give rise to the social value of trust (6.2.4), but trust is not yet political. It becomes political when it becomes contested – when people disagree about who or what can be trusted. From a virtue-ethical perspective, such disagreement reflects differences in the reasons and narratives through which people have arrived at their particular conceptions of trust. Therefore, rather than deciding which conception of trust is correct, we should examine the commitments and beliefs that underlie each conception.

On this view, we see that there can be broad agreement that trust is valuable (i.e., agreement on the concept) but disagreement about what it means or what problem trust is meant to resolve (i.e., disagreement on the conception). The difference is therefore understood to originate in the different configurations of virtues, i.e., the same capacity can narrate different dispositions, and different configurations of capacities can yield different meanings for a single value, which is why trust founded on sincerity differs from trust founded on accuracy.

Ontologically, this account takes human beings to be primarily creative beings (as opposed to, e.g., competitive). The drive and joy of living arise through engagement with resistance, since resistance is what awakens and engages creativity. Living well therefore consists not in enduring suffering or striving towards happiness, but in navigating social complexity through meaning-making and creating new meaning for oneself. The ethos of creativity thus stands opposed to what we could call the ethos of production with its orientation

towards rationality, solutionism, and performativity. On the account of creativity, valuation is not primarily about reaching a good/bad verdict, but about asking whether, and in what way, something is meaningful to the valuer.

On this view, the social order is open-ended, and its building blocks are considerations rather than rules (6.6.5). A rule is basically an answer to an unasked question; thus it can be followed or rejected, obscuring the question that prompted the rule in the first place. A consideration, by contrast, says what is to be attended to while leaving the answer open, thereby preserving the space for valuation. A consideration cannot be breached or sanctioned in the same way as a rule; it can only be attended to in different ways.

I have argued that foundational value commitments that manifest as irreconcilable disagreements diverge at the level of problematisation rather than at the level of the verdict (6.6.2), which helps to explain why such disagreements can appear irreconcilable, since attempts to settle them target the wrong level of thought. Two parties may reject the same solution for entirely different reasons, or accept the same solution while conceptualising the underlying problem differently. Settling the disagreement at the level of the solution can therefore make the apparent agreement performative or illusory. Thus, truly understanding value disagreement requires looking beyond proposed solutions to how the problem itself is framed.

Problematisation at the individual level and solution-seeking at the social level thus connect the individual and social dimensions of the virtue-ethical perspective. An individual disposition that is not shared by others generates a capacity for social virtue, but not yet a social value; it becomes socially meaningful only through relations with others and becomes political when it becomes an object of contestation. Virtue-ethical politics therefore has to attend to the diversity of social values. At the core of this diversity lies what moralists, realists, and agonists identify as a key feature of politics, namely power, understood in terms of positionality relative to others. On these accounts, power structures the social order, and political questions are consequently understood as questions about power, either primarily, as in realism and Mouffe, or derivatively, as in Rawls, Honig, and Connolly (6.5).

By contrast, virtue-ethical politics is concerned with empowerment, and the purpose of politics is accordingly to secure the conditions under which people can engage in valuation, since valuation is itself a source of empowerment. What such conditions must secure, building on Miyasaki (2022a), is what I call constitutive resistance (6.5), understood as engageable obstacles, generated by the diversity of social perspectives, that can be overcome without being eliminated. Diversity is thus constitutive of resistance, and resistance is in turn constitutive of valuation,

since without something that pushes back there is nothing to value, only something to accept or reject.

Thus, maintaining this resistance requires keeping solutions open to disagreement and contestation, which, in turn, leads to a different view of social stability. An order based on clear and enforceable rules may appear stable, but this stability can be fragile if people follow those rules without thinking about why they matter. In that case, their commitment depends on external authority rather than on their own ability to create meaning and understand why the rules are worth following. So, a society that eliminates disagreement also eliminates the resistance that gives people something to value. Consequently, friction can make a social order more resilient rather than simply less stable. While agonists have shown that contestation (and thus, friction) is central to political life, the virtue-ethical account tells us that it is also necessary for its resilience, i.e., sustainable stability (6.6.4).

Taken together, these commitments describe what I have called self-creative democracy (6.6), that is, a politics concerned with sustaining the conditions for valuation, which moves beyond cyclical but formally prescribed procedures characterising self-corrective democracy. Self-creative democracy instead creates itself anew in the sense that its citizens continually create new meanings for themselves and, through this process, participate in creating the social and political order. The political and social order is therefore understood as a space for meaning-making in the public sphere, rather than primarily as a structure of rules or principles that discipline it. While self-correction seeks to restore a standard, self-creation considers standards as objects of valuation and reconsideration.

Virtue-ethical politics demands of citizens that they not be politically indifferent, but it does not require agreement on much beyond treating the question of how we are towards each other as a primary question of social life. That said, even this seemingly thin demand rests on the ontological commitment set out above, namely that human beings are creative, possess the drive to engage with resistance that is roughly equal to their capacities, and care about the conditions under which others, and not only they themselves, are able to value and create. Moreover, shifting from a rules-based order to a consideration-based one entails that instead of rules doing the heavy lifting of creating social order, the responsibility shifts to people engaging with it proactively through the meaning-making process of valuation, alongside honesty and sincerity in expressing one's actually held commitments, no matter how alien or uncomfortable they may seem in the particular society, rather than having to perform for instrumental or otherwise non-intrinsic reasons.

6.7.2 Objections

The virtue-ethical ethos I have developed here declines to specify what substantive aims politics ought to achieve. Such a view is likely to invite scepticism, which I will now address in the hope of providing greater clarity about its rationale and underlying commitments. Below, I consider seven objections. If my replies do not satisfy a critical reader, further objections would be most welcome, because, as with any attempt to introduce new perspectives, this account requires sustained scrutiny if it is to be refined and improved over time. At this point, I think I can answer six objections, but one I cannot, at least not comprehensively, and I will admit as much when we reach it.

1.

So, let us begin with the most basic question: what does this ethos actually ask of us? An ethos, whatever else it does, grounds value commitments in itself, but my account specifies no institutions, no principles, and no conception of a just, equal, or good society. So either it has no commitments at all, or it relies on commitments it has not owned up to. To reply briefly, the ethos does have value commitments, but these do not specify the substantive content of the good. Rather, they provide two foundational value considerations that I take to be characteristic of the human condition.

First, it is the assumption that social life is sustained by a shared sensibility that ethical considerations matter to human beings in the first place, and that the question of ‘how we are towards each other?’ guides that sensibility (see 1.4.1 and 6.3.1). Thus, the virtue-ethical ethos considers humans to care about their relations with others within the social sphere. However, it is not primarily about caring for others *per se*, but about caring about one’s own sense of and feelings towards others.

Second, following from the assumption that humans are first and foremost creative beings (6.4), as opposed to, e.g., competitive beings, avoiding harm becomes an overarching common value because creativity and constitutive resistance depend on it (6.6.5), and because harm (and domination alike, for that matter) reduces the possibility of roughly equal resistance in the Nietzschean sense (Miyasaki 2022a, 93 fn. 40). Thus, the reason for avoiding harm is neither altruistic nor strategic but affectively grounded, since we care about how we stand towards others, including those who oppose us, and this care is bound up with the way others’ capacity to resist sustains our own creativity.

2.

These two commitments form an admittedly thin core of the virtue-ethical view, which in turn invites the next objection. Let us suppose someone grants both commitments (that ethics matters and that harm is to be avoided). I have argued that politics should not be pursued for its own sake, yet I also hold that virtue-ethical politics provides the conditions for valuation rather than producing solutions, and that the good lies in valuation itself. How, then, is valuation for its own sake any different from politics pursued for its own sake? This question seems to make sense if we agree with Rawls (1996) and Williams (2005) that politics must be kept apart from comprehensive doctrines because if these, being foundationally irreconcilable, enter the political sphere, political and social stability will be impossible, and without stability, in turn, just or otherwise good order will not be attainable either. Thus, to reject stability as the primary aim of politics might seem like a refusal to take the problem seriously.

I think the objection misdescribes valuation, treating it as something one engages in for an instrumental end, in the way one takes up a project. But valuation is not chosen; it is prompted by attending to situations, circumstances, and oneself. The process of valuation through engagement with resistance sits at the core of the human way of being, or more precisely, of becoming (5.2.2). Consider Murdoch's M&D example ([1970] 2014, 16–24; discussed in 1.4.2 and 6.3.2). M does not decide to reconsider her disposition towards D. Something about D initially feels alien and uncomfortable, and M therefore experiences hostility towards her. She encounters a form of resistance that she cannot simply ignore, and the valuation that follows leads her to recognise that her view of D may be clouded by her own qualities rather than by D's. The impetus is not external pressure or moral instruction, but M's own epistemic doubt. Being internally driven, it was not something she could have declined.

Consider also Nietzsche's claim about resistance: "A living being wants above all else to release its strength; life itself is the will to power, and self-preservation is only one of its indirect and most frequent consequences" ([1886] 1998, §13, original emphasis). Read together, the Murdochian process of valuation is an expression of Nietzsche's will to power, and if both are right, valuation is not something that politics, or anything else, assigns to us.

By contrast, the purpose of virtue-ethical politics is to provide the conditions for social-diversity-driven political contestation (6.6.2), and thus to enable precisely the kind of resistance through which human beings can *release their strength*, in the Nietzschean sense. Politics therefore has an instrumental aim, even though the activity it enables does not, and the parallel between valuation and political action fails.

The same cannot be said of agonism, though. Promoting contestation for its own sake creates a problem, as it requires people to already have the very disposition that contestation is supposed to develop. For example, meaningful adversarial sparring (4.2.3) requires a willingness to take the opposing commitments as seriously as one's own commitments. Someone who contests simply for the sake of contesting has little reason to pay attention to what they are actually engaging with. Agonists therefore rely on a disposition that they also claim contestation itself will produce (4.6), creating a difficulty that I return to below.

My view stands closest to Connolly and Honig among the agonists, but it places greater weight on the process of ethical valuation and on the political agent as an artist. It tasks politics with creating the conditions for roughly equal resistance and locates the good in the activity of valuing itself (6.5). The good it identifies therefore belongs to social life and its values, rather than to the practice of politics considered separate from them.

3.

I elevate creativity as the primary epistemic authority (6.4). Why is that any less moralistic than elevating Rawlsian moral powers (capacity for the sense of justice and the good, see 2.5.1)? And what stops a creativity-centred ethos from hardening into a dominating organising principle of its own?

The objection assumes that an emphasis on creativity must operate as a principle. But self-creative order relies on non-domination, and its organising ethos is a consideration rather than a rule, since its aim is not harmonisation or stability for their own sake but roughly equal resistance, and it coordinates through consideration rather than strict supervision or surveillance (6.6.4–6.6.5). What creativity thereby avoids is precisely a rule-based stability that is at once rigid and fragile, two features that make an order lean towards domination.

4.

Next, the claim that politics should not aim at particular solutions raises the question of how grave harm is avoided on my account. We can consider this from two perspectives. The first concerns what an open-ended and low-surveillance politics allows to happen, for if political solutions remain open-ended, and if people are understood to enjoy resistance and to embrace the indeterminacy and unknowability that come with it, then I seem to be asking them to embrace whatever fate arrives, famine included. The second concerns what an uncoerced creativity permits people to do: if there is no structural mechanism of coercive constraint and no Rawlsian public reason, what prevents creativity from being appropriated by those most capable of inventing new forms of oppression? Taken

together, the charge is that an account too thin on aims and constraints cannot adequately account for grave harm, whether suffered or inflicted.

To reply, we ought to distinguish between stability as a political aim and the stability of substantive arrangements or conditions (6.6.4). What I have argued against is treating stability as the primary good of political life, and keeping its aims open does not imply indifference to human suffering. The substantive reply is that grave harm is ruled out by what creativity is, in the sense developed here, namely a process of meaning-making through engagement with resistance (6.4). In this sense, swimming with the current (uncreatively) does not prompt creative movement through water. The resistance of the current must be engageable, so that the swimmer can meet it without drowning or simply drifting. Starvation is an obstacle no person can take on, as it consumes the very capacity that engagement would draw upon. Genocide and torture likewise eliminate the capacity to resist and thereby destroy the condition under which engagement is possible at all.

So, Nietzsche does not recommend enjoying suffering. My account goes further by adding an explicit social dimension, since caring about *how we are towards each other* (1.4.1) means that another person's capacity to resist becomes a concern of mine, rather than merely a fact about them. In the ordinary sense of ingenuity, the kind of appropriation imagined by the critic therefore remains possible, but only in the case of individuals incapable of ethical valuation (6.3.1). Throughout, however, I have assumed that human beings are creative agents who also care about others' capacity to engage in constitutive resistance. This is because, on the view developed here, such engagement can itself generate joy through the experience of empowerment (6.6.2).

To reiterate, my account does not claim that famine or atrocity is impossible in practice. Rather, it claims that when political processes fail to create the conditions for roughly equal, resistance-induced valuation, and when socially generated harm is present, individuals will not experience empowerment but suffering instead. This is because power that depends on harm and suffering ultimately undermines itself, becoming trapped in an endless cycle of quantitative power-seeking (Miyasaki 2022a, 77–88; 5.2.1). This is not indifference to harm; quite the contrary, it is a claim about the cost of conducting politics without an ethos of virtue-ethical valuation.

5.

This brings me to the objection I do not think I can answer sufficiently well: if constitutive resistance requires that agents remain within the limits of

engageable contestation, then how is it determined where those limits lie? Who decides, and how?

Since the limits of constitutive resistance must themselves remain contestable, the only check against their abuse is the one already discussed, namely that domination destroys the resistance on which the empowering feeling of power depends (6.5; cf. Miyasaki 2022a, 94). But whether this check is sufficient requires further consideration. That said, self-creative order does not reject institutionalised apparatuses or rules altogether; it merely supplies a criterion for their use, in that legislation, courts, and other institutions support the self-creative ethos insofar as they sustain the conditions for valuation.

6.

Much of the argument so far has engaged with the individual rather than with social interaction (5.2.2), so it may not be immediately obvious to a reader how one moves from that figure to the social phenomenon of politics, or what a creativity-centred social policy would look like.

I illustrated this in 6.3.3 through the case of Dame Cicely Saunders's hospice movement, which shows how an individual way of thinking and acting, when shared with others, can become a social virtue. In this case, what began with one person's changed view of dying developed into a shared practice that others could join, learn from, and develop further. So, once others take it up in practice, it no longer belongs to any one person; it gains wider support and can eventually become an organised practice supported by the state.

On the virtue-ethical view, politics is not limited to formal or institutional settings. Whenever people challenge existing beliefs or arrangements, the interaction becomes political. This puts the individual, rather than the system or structure, at the centre of the ethos, because the virtue-ethical view considers politics as something that does not abstract away from people or reduce political engagement to procedures, but as working towards ensuring conditions for people to engage and contest. My account therefore provides an ethical framework for engaging with substantive questions, e.g., what a just society should look like, without providing answers to those questions.

7.

And now the last objection. Even if the virtue-ethical ethos has traction and it survives these critical comments, perhaps none of this amounts to an actual alternative to moralism, realism, and agonism? Perhaps it is merely an ethos of theorising that any of them could accommodate?

To answer this, we must look at what each approach requires of political agents. Rawls requires individuals to possess the two moral powers (1985, 233; see 2.5.1) and to accept the constraints of public reason (1996, 212–254), excluding proponents of unreasonable doctrines (1996, 39) in order to reach a consensus. But an ethos of creative valuation conflicts with that consensus-seeking orientation because, on the ontological commitment that human beings are animated by challenge rather than by stability, consensus-seeking is undesirable however benevolent it looks. The same holds for realists, who seek legitimate stability (Williams 2005, 4–9). So, ultimately, on my account, politics is necessary for creating a field of engagement, not for achieving some end (6.5).

Agonists fare far better in terms of adopting an ethos of creative valuation and its underlying question of ‘how we are towards each other?’, though only Honig (4.2.3) and Connolly (4.3.3). Mouffe seeks a counter-hegemony whose emancipatory content is situated within a particular socio-historical context (2013, 14; 4.4.1), and such a hegemony, benevolent as it might be, is not compatible with my account.

So, can ethical creativity and an emphasis on valuation be incorporated into Honig’s or Connolly’s accounts?

Honig’s argument is that every settlement produces remainders ([1993] 2023, 5), locating responsibility institutionally rather than in the agent (4.6). She therefore sees the impulse to contest as emerging from political contestation itself. Her focus is on continuously producing disruption, “not for contest’s sake but for the sake of those remaindered by even the best, most tolerant, democratic political orders” (ibid., x). This makes her account explicitly emancipatory. Although her “method” (ibid.) – like mine – draws on Nietzsche’s idea of resistance, Honig sees politics as a movement between two poles: “[p]olitics consists of practices of settlement and unsettlement, of disruption and administration, of extraordinary events or foundings and mundane maintenances” (ibid., 205). If either side is abandoned, she argues, politics becomes depoliticised and produces remainders. Her aim, then, is specifically to give voice to those remainders, with disruption bringing those who have been depoliticised back into the political arena, while remaining committed to “democratic equality” (ibid., xi).

My account is substantively much less ambitious. Politics is concerned only with securing the conditions for resistance-driven creative valuation. Yes, it does not work towards equality, but this is not because it is opposed to equality. Rather, even relatively thin concepts such as ‘democratic equality’, ‘remainders’, and ‘displacement’ presuppose a particular conception of the political end to be pursued, i.e., they carry some substantive content, even if that content is not fully specified. Even if Honig’s agonism is understood as engaging in “sparring”,

sparring is not undertaken simply for its own sake; it serves the broader aim of preserving or restoring equality.

If my account, which considers creative valuation intrinsically valuable, were incorporated into Honig's framework, the two would therefore come into conflict. At some point, we would need to know how to identify remainders and displacement. Self-creative democracy, by contrast, is not primarily about disruption but about resilience, and it places greater emphasis on keeping institutionalised monitoring and discipline to a minimum. In this sense, Honig's disruptive agonism can look like a struggle against something, thereby presuming that there is a pre-politically determined 'good' and 'bad' against which disruption is directed. Moreover, to disrupt something, that something must first be identified substantively as a remainder before disruption can begin. My account takes a different approach. It does not begin with these evaluative categories, but with valuation itself, that is, with the process through which individuals engage with what is intelligible within their own lifeworld. The struggle, understood as engagement with resistance, begins with the individual and only subsequently takes on a social and political form.

Now to Connolly. There are strong parallels with his work, which is not surprising, since we both draw heavily on Nietzsche. Like mine, Connolly's account locates agonism in the dispositions of individuals rather than in institutional design. It diagnoses resentment as grievance over unrealised possibilities ([1991] 2002, 191), which stands close to the quantitative power-seeking I describe in 5.2.1 (Miyasaki 2022a, 77–88). It asks that agents be able to "eventually modify what they are, in form or content or both" (Connolly 1995, 184). My account likewise is about modifying oneself, not the system, since I take modifying the system to inevitably require exercising some form of domination. And critical responsiveness runs very close to valuation, since Connolly's combination of anticipation, critical engagement, and self-revision is a near neighbour of the attention I have borrowed from Murdoch, and on both accounts the disposition is triggered by a difference that creates resistance.

Connolly's account of political engagement aims to remain open to identities as they emerge and change, breaking the barriers that existing cultural-moral conventions about identity have set up (1995, 185). His aim, in other words, is to catch a social process early. Such pluralisation is, in my terms, a social value, so that critical responsiveness is a way of coming to value what was not valued before. This reads close to Murdoch's ([1970] 2014, 16–24) M&D example, where M begins by seeing D as merely the (unpleasant) 'other', and after something like critical responsiveness comes to see her as differently constituted rather than deficient.

So how, exactly, is my account different from Connolly's? In short, his approach tells us how to respond well to values, whereas mine tells us how values come to be and have meaning at all. Consider trust. Connolly can tell us what it is to be responsive to those who trust differently. What he does not explain is why trust matters to us in the first place, why it matters differently to different people, or how two people who both sincerely claim to trust one another can act in ways the other experiences as betrayal. While Connolly would likewise reject the view that trust is a value held in common and then applied unevenly, he merely takes this as given, whereas the virtue-ethical view asks how its meaning was formed through valuation. What is ultimately understood as a plural society is therefore not a plurality of commitments but a plurality of reasons behind those commitments.

So, the two accounts are largely complementary, and an ethos of creative valuation can underlie Connolly's agonism by supplying the account of formation that Connolly's ethos of critical responsiveness assumes. But they do not fit in every respect. I will outline two core differences, namely the aim of politics and the conditions/requirements.

First, Connolly's target is the reduction of existential resentment towards the different other, grounded in the dogmatisation of one's own identity ([1991] 2002, 120–121). My account does not treat resentment as something to be reduced per se, but as something whose sources need to be understood from the perspective of those who are viewed as resentful, considering its emergence from situations where resistance is grossly unequal. What therefore needs addressing is the inequality rather than the affect. In this respect, Connolly's project lists resentments that have already been conceived as emerging from past injustices, e.g., the feminist and gay rights movements (ibid., 85), but this does not tell us anything about how to anticipate and respond to future processes. Connolly's theory therefore seems to explain the past and is able to provide normative insight insofar as the future is not radically different from the past. Within the ethos of creative valuation, by contrast, resentment is itself a form of legitimate resistance, and the account directs us to engage with it at the level of reasons rather than commitments, so as to increase intelligibility between the parties (6.6.2), even if they seem radically apart in their commitments and conduct.

Second, Connolly acknowledges that democratic virtue already assumes a democratic way of life. He asks us to imagine that democracy and its virtues have developed together ([1991] 2002, 193), and while this is a legitimate move for his purposes, it means that his account begins with a situation in which a rough equality of resistance already exists. My account cannot begin there, because I take politics to be partly about creating those conditions in the first place (6.6.2).

This is also where his account differs from the Nietzschean will to power, as those for whom resistance is neither challenging nor intelligible fall outside the conditions his account assumes. Further, because Connolly assumes these conditions, he does not need to explain how an individual moves from where they are to the kind of disposition his ethos requires. He shows us what that disposition looks like and describes some of the practices through which it can be developed, but he does not explain what starts this process in someone who is not already inclined towards it. Creative valuation does address this, arguing that meaning-making through engagement with differences that create resistance is not something we simply choose (or are forced) to do. The process begins when something we encounter moves us, and meaning is created through that engagement whether or not we intended to seek it.

To conclude, these replies certainly do not amount to a complete defence, though I doubt any set of replies could. We can nevertheless take at least three things from this engagement with possible critics. First, the two commitments the ethos of creative valuation requires of a society – that ethics matters and harm is to be avoided – are thin, but they are commitments nevertheless, and unlike agonism the account does not presuppose a disposition that the practice it licenses is itself meant to produce. Second, an ethos, whether the one proposed here or some other, is a necessary dimension of theorising, either on its own or as an underlying layer beneath more practically oriented discussion, insofar as values, meaning-making, and lived experience are what make us human. Third, putting the process of valuation before values – asking ‘what is relevant here?’ and ‘what am I not seeing?’ rather than, for example, ‘how should I be honest?’ – reminds us of what animates social creativity.

CONCLUSION

This study argued that political philosophy tends to treat values as they are, as commitments already given and verdicts already reached, and then asks what politics should do when they collide. And when it does examine how values are formed, it tends to treat their formation in terms of linear causation. Treating them as verdicts implies that values are considered as offering solutions, and treating them as such at face value obscures the variety of meanings that a value can hold.

The social value of trust, for example, can mean quite different things depending on the virtues that constitute it. Trust grounded in sincerity asks that what a person says be what they believe, while trust grounded in accuracy asks that a person take the trouble to get things right, whatever the state of their beliefs (cf. Williams 2002). We can therefore agree that trust is valuable while disagreeing about what counts as a betrayal of trust.

The disagreement, then, is not about whether to value trust; it may instead concern what is being valued in the first place. If we look only at the level of the verdict, the conflict may appear to be a clash between irreconcilable commitments about the true value of trust. To see past this, we need to look closely at the process of valuation itself, and at how it creates the meaning that gives a value its content. On this view, what a value is depends on how its meaning is created.

I discussed three approaches to normative theory that do not ask how values come to have meaning. I argue that to understand disagreement, we need to look beyond disagreements about answers or solutions and ask whether the disagreement goes deeper, that is, whether people understand the problem itself in different ways. Attending to problems is, after all, a driving force of human life. Throughout the study, I show how engaging with problems, understood as forms of resistance, can animate human creativity. Trust, in this sense, can be understood as a solution framed as a social value, but also as a problem in its own right. This captures a paradox of creative valuation: every verdict that resolves a problem also reframes it, generating new forms of resistance and friction that sustain the creative and political life of society.

Normative theorising is therefore concerned with what kind of political arrangement is desirable and why. Any positive normative theory constructs its political arrangements in terms of values while assuming certain value configurations as their basis. A theory therefore needs to be explicit about the meaning of the values on which it relies. The same applies when a theory considers a particular political arrangement. If, for instance, a normative theory

appeals to trust, we need to understand what trust means within that theory, and this requires understanding the process of valuation through which that meaning is formed. That process of valuation is embedded in the ethos of theorising. This study explores three approaches to normative theorising and their relationship with values, primarily in the context of moral values, and while each has an implicit ethos (as all normative theories do), none discusses this ethos explicitly.

Given that I take ethics as the guiding mode of thinking and the guiding question of thinking about the social sphere as ‘how we are towards each other?’, I consider ethics of being, that is, virtue ethics, as the foundational value framework for our social relations. On this grounding, I introduce virtue-ethical creative valuation as the way to approach the process of how we create meaning for values. A comprehensive summary of this is laid out in 6.7.1, so I will not repeat it here. I want only to emphasise that I understand virtues not as expressions of already established values, as traditional virtue-ethical theories do, but as epistemic capacities through which values acquire meaning.

So, I propose considering an ethos of theorising as a distinct dimension of normative theorising, one that precedes its empirically more concrete dimensions. And since ethos is about a configuration of value commitments that a theory adopts, and since I understand humans as value creators, I turn to virtue ethics as an account of human traits as capacities for ethical meaning-making. This view differs from moral theories that seek to determine what the good life, the best action, or an adequate solution is. I argue that such approaches are less concerned with creating meaning than with establishing a particular order, however benevolent that order may be.

I also emphasise creativity as the process of meaning-making in normative thought. As I have argued throughout this study, creativity animates human life, and I propose that it can likewise animate political philosophy. We might therefore approach political theorising as an activity of creating the meaning of value through politics, much as artists create the meaning of value through art.

Creative political philosophy would not seek to predict the social world or to derive a solution from principles, but to create new ways of seeing what matters and new ways for humans to relate to one another within political settings. The political philosopher’s task, consequently, is not primarily to ask, ‘What should we do to live well together?’ but to ask the prior question: ‘What might we come to value, and what new forms of social life might become possible if we attended, looked, and saw differently?’

The art of political philosophy, in this sense, lies in facilitating the exercise of capacities that make new meaning through valuation possible. Creativity is

understood both as an activity and as a virtue, as it sustains the capacity to generate new meanings. Political philosophy becomes creative, then, not by abandoning ethical constraint, but by recognising that creating new ethical possibilities is itself part of the work of political thought. One such ethos is developed here through the idea of self-creative democracy. It begins from the thought that the self of democracy is the self of each of its members, so that what democracy creates and re-creates is not primarily its procedures but the meanings by which its members live. Where self-corrective democracy restores a standard, self-creative democracy continually challenges and re-creates its own. Because the virtue-ethical account treats valuation as the process of meaning-making, and meaning-making as the primary source of political value, such an order secures the conditions for valuation rather than pursuing any substantive or structural end. Its aim is to sustain constitutive resistance, the engageable and roughly equal difference without which there is nothing to value; it frames its arrangements as considerations, which say what is relevant, rather than as rules, which provide ready-made answers; and it coordinates rather than supervises. This does not mean that there are no rules whatsoever; rather, instead of rules doing the heavy lifting of creating social order, the responsibility shifts to people to engage with it proactively. The meaning-making process, alongside Nietzschean honesty in facing hard truths and sincerity in expressing one's actually held commitments rather than performing them, is what is required of people. Such an order therefore keeps its standards open to contestation and re-making, and so generates healthy friction, which, in turn, is what makes it resilient. Orders built on correction and rule-following may seem sturdier, but their rigidity also makes them fragile.

This study has covered a lot of ground in thinking about ethical values and politics, and its scope has therefore been deliberately wide and exploratory. If I am right that ethical valuation and creativity are what animate human life, and that politics should address value diversity at its roots rather than mistake solutions for problems and vice versa, then much remains to be explored, both conceptually and in more substantive theorising about what a virtue-ethically driven politics might look like, for example, how representation might work within a self-creative democracy that treats problems, rather than fixed solutions, as a source of constant renewal.

The ideas developed here open multiple avenues for further research. The most immediate concerns the institutional implications of an ethos of creative valuation. For instance, in political representation, traditional accounts are largely concerned with what is represented, e.g., identities, interests, or ideas, and with translating these into political power. But if valuation and creativity are understood as activities (as they are in this study), rather than values being taken

as settled positions and thus as the elements of inquiry, and if the question of power is deferred to the empowerment through constitutive resistance, the object of representation also changes. In this case, representation would not be accountable in its decision-making to what people value, but to the conditions under which they can create and revise the meanings of values for themselves, shifting the question, in Murdochian terms, from what is being seen to how seeing takes place. The task would not be to aggregate preferences or impose a particular conception of value and thus create power, but to create possibilities for different ways of seeing, while ensuring that any decisions made foster further resistance. Ostrom's work provides one possible starting point for developing such an approach by directing attention to the capacities and practices through which institutional arrangements are sustained (see 6.6.5), but her account does not envisage deliberately creating further resistance. On the virtue-ethical account, such resistance could be understood as deliberately creating an environment of friction in which failure is accepted not as inevitable, but as a positive building block. The distinction between coordinative and supervisory order raises a related question. Creative valuation cannot imply the absence of rules altogether, as some activities do require hard constraints, and it must be determined under which conditions coordination should give way to rules.

The demands this ethos places on citizens (see 6.7.1) also require further examination, raising questions about how such capacities are formed, and about the extent to which they can reasonably be expected of citizens. Further, the argument is developed largely with the individual in view, but how constitutive resistance would work in large, anonymous, and heavily mediated settings – institutions, mediatised public spheres, and technological environments – still needs to be tested.

A further avenue concerns the conditions under which an ethos of self-creative democracy can emerge, opening a door to critical theorising: rather than asking only which norms or institutions produce particular outcomes, we could investigate the means by which constitutive resistance is deliberately made unavailable, including fostering indifference (resembling the Nietzschean Last Man), ideological closure (as manifest in polarisation), and the manipulation of informational environments (as visible in the so-called post-truth phenomenon), and identify the processes through which the capacity to create meaning becomes narrowed.

And finally, my thanks to the reader for coming this far. On an account that treats resistance as the source of renewal and empowerment, the best I can wish for this study is that it will be argued with.

REFERENCES

- Abbas, Syed Zamanat. 2023. 'The Veil: A Silhouette of Autonomy and Empowerment'. *Culture & Psychology* 29 (2): 247–259. <https://doi.org/10.1177/1354067X221115852>.
- Abbey, Ruth, ed. 2015. *Feminist Interpretations of John Rawls*. Penn State University Press. <https://doi.org/10.1515/9780271063133>.
- Allen, Jonathan. 2001. 'The Place of Negative Morality in Political Theory'. *Political Theory* 29 (3): 337–363. <https://doi.org/10.1177/0090591701029003002>.
- Angier, Tom. 2016. *Either Kierkegaard/or Nietzsche: Moral Philosophy in a New Key*. First issued in paperback. *Intersections Continental and Analytic Philosophy*. Routledge, Taylor & Francis Group.
- Annas, Julia. 1992. 'Ancient Ethics and Modern Morality'. *Philosophical Perspectives* 6: 119. <https://doi.org/10.2307/2214241>.
- Annas, Julia. 2011. *Intelligent Virtue*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199228782.001.0001>.
- Anscombe, G. E. M. 1958. 'Modern Moral Philosophy'. *Philosophy* 33 (124): 1–19. <https://doi.org/10.1017/S0031819100037943>.
- Ansell-Pearson, Keith. 2007. 'Introduction: On Nietzsche's Critique of Morality'. In *On the Genealogy of Morality*, revised ed., by Friedrich Nietzsche. *Cambridge Texts in the History of Political Thought*. Cambridge university press.
- Arendt, Hannah. (1958) 1998. *The Human Condition*. With Margaret Canovan. 2. ed., Repr. Univ. of Chicago Press.
- Aristotle. 2014. *Nicomachean Ethics*. Translated by Charles David Chaneel Reeve. Hackett Pub. Co.
- Aristotle. 2017. *Politics: A New Translation*. Translated by Charles David Chaneel Reeve. The New Hackett Aristotle 6557. Hackett Pub. Co.
- Ashford, Elizabeth, and Tim Mulgan. 2018. 'Contractualism'. In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. <https://plato.stanford.edu/archives/sum2018/entries/contractualism/>.
- Aytac, Ugur. 2022. 'Political Realism and Epistemic Constraints'. *Social Theory and Practice* 48 (1): 1–27. <https://doi.org/10.5840/soctheorpract2021129145>.
- Aytac, Ugur, and Enzo Rossi. 2023. 'Ideology Critique without Morality: A Radical Realist Approach'. *American Political Science Review* 117 (4): 1215–1227. <https://doi.org/10.1017/S0003055422001216>.
- Azzano, Lorenzo, and Andrea Raimondi. 2023. 'Vices, Virtues, and Dispositions'. *TheoLogica: An International Journal for Philosophy of Religion and Philosophical Theology* 7 (2). <https://doi.org/10.14428/thl.v7i2.67873>.
- Baderin, Alice. 2021. 'The Continuity of Ethics and Political Theory'. *The Journal of Politics* 83 (4): 1734–1746. <https://doi.org/10.1086/715068>.

- Baehr, Jason. 2015. 'The Four Dimensions of an Intellectual Virtue'. In *Moral and Intellectual Virtues in Western and Chinese Philosophy: The Turn toward Virtue*, edited by Chienkuo Mi, Michael Slote, and Ernest Sosa. Routledge, Taylor & Francis Group.
- Bagg, Samuel. 2022. 'Realism against Legitimacy: For a Radical, Action-Oriented Political Realism'. *Social Theory and Practice* 48 (1): 29–60. <https://doi.org/10.5840/soctheorpract2021129146>.
- Baker, Lynne Rudder. 2015. 'Human Persons as Social Entities'. *Journal of Social Ontology* 1 (1): 77–87. <https://doi.org/10.1515/jso-2014-0037>.
- Beckstein, Martin. 2011. *The Dissociative and Polemical Political: Chantal Mouffe and the Intellectual Heritage of Carl Schmitt*. February. <https://doi.org/10.5167/UZH-50536>.
- Bhaskar, Roy. 2008. *Dialectic: The Pulse of Freedom*. Classical Texts in Critical Realism. Routledge.
- Blum, Lawrence A. 2023. 'Iris Murdoch'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Nodelman, Uri. <https://plato.stanford.edu/archives/fall2023/entries/murdoch/>.
- Brobjer, Thomas H. 2003. 'Nietzsche's Affirmative Morality: An Ethics of Virtue'. *The Journal of Nietzsche Studies* 26 (1): 64–78. <https://doi.org/10.1353/nie.2003.0020>.
- Brobjer, Thomas H. 2025. *Nietzsche's Revaluation of All Values*. 1st edn. Cambridge University Press. <https://doi.org/10.1017/9781009421652>.
- Brubaker, Rogers. (1984) 2010. *The Limits of Rationality: An Essay on the Social and Moral Thought of Max Weber*. Routledge Library Editions: Weber. First issued in paperback. Routledge.
- Buchanan, Allen, and Russell Powell. 2017. 'De-Moralization as Emancipation: Liberty, Progress, and the Evolution of Invalid Moral Norms'. *Social Philosophy and Policy* 34 (2): 108–135. <https://doi.org/10.1017/S0265052517000231>.
- Burelli, Carlo. 2022. 'Political Normativity and the Functional Autonomy of Politics'. *European Journal of Political Theory* 21 (4): 627–649. <https://doi.org/10.1177/1474885120918500>.
- Burelli, Carlo, and Chiara Destri. 2022. 'The Sources of Political Normativity: The Case for Instrumental and Epistemic Normativity in Political Realism'. *Ethical Theory and Moral Practice* 25 (3): 397–413. <https://doi.org/10.1007/s10677-021-10243-y>.
- Caprioglio Panizza, Silvia. 2022. 'Attention'. In *The Murdochian Mind*, by Silvia Caprioglio Panizza and Mark Hopwood. Routledge Philosophical Minds. Routledge, Taylor & Francis group.
- Chahboun, Naima. 2019. 'Ideal Theory and Action-Guidance: Why We Still Disagree'. *Social Theory and Practice* 45 (4): 549–578. <https://doi.org/10.5840/soctheorpract2019123071>.
- Chambers, Clare. 2017. 'II—Ideology and Normativity'. *Aristotelian Society Supplementary Volume* 91 (1): 175–195. <https://doi.org/10.1093/arisup/akx008>.
- Clark, Rebecca L. 2024. 'Moderate Realist Ideology Critique'. *European Journal of Philosophy* 32 (1): 260–273. <https://doi.org/10.1111/ejop.12929>.

- Coady, C. A. J. 2024. 'The Problem of Dirty Hands'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Nodelman, Uri. <https://plato.stanford.edu/archives/spr2024/entries/dirty-hands/>.
- Cohen, G. A. 1989. 'On the Currency of Egalitarian Justice'. *Ethics* 99 (4): 906–944.
- Cohen, G. A. 2003. 'Facts and Principles'. *Philosophy & Public Affairs* 31 (3): 211–245. <https://doi.org/10.1111/j.1088-4963.2003.00211.x>.
- Cohen, G. A. 2008. *Rescuing Justice and Equality*. Harvard University Press.
- Collier, Andrew. 1999. *Being and Worth. Critical Realism: Interventions*. Routledge.
- Connolly, William E. 1994. *Political Theory and Modernity*. 2. print. Cornell Paperbacks. Cornell Univ. Press.
- Connolly, William E. 1995. *The Ethos of Pluralization*. Borderlines, v. 1. University of Minnesota Press.
- Connolly, William E. (1991) 2002. *Identity/Difference: Democratic Negotiations of Political Paradox*. Political Theory. Expanded ed., [Nachdr.]. Univ. of Minnesota Press.
- Connolly, William E. (1993) 2002. *The Augustinian Imperative: A Reflection on the Politics of Morality*. Modernity and Political Thought. New edition. Rowman & Littlefield.
- Connolly, William E. 2005. *Pluralism*. Duke University Press. <https://doi.org/10.1215/9780822387084>.
- Connolly, William E. 2011. *A World of Becoming*. A John Hope Franklin Center Book. Duke University Press.
- Cozzaglio, Ilaria, and Greta Favara. 2022. 'Feasibility beyond Non-Ideal Theory: A Realist Proposal'. *Ethical Theory and Moral Practice* 25 (3): 417–432. <https://doi.org/10.1007/s10677-021-10206-3>.
- Cross, Ben. 2020. 'Radicalizing Realist Legitimacy'. *Philosophy & Social Criticism* 46 (4): 369–389. <https://doi.org/10.1177/0191453719857129>.
- Cross, Ben. 2022. 'How Radical Is Radical Realism?' *European Journal of Philosophy* 30 (3): 1110–1124. <https://doi.org/10.1111/ejop.12710>.
- Cross, Ben. 2023. 'Realist Legitimacy: What Kind of Internalism?' *Philosophy & Social Criticism*, November 3, 01914537231211030. <https://doi.org/10.1177/01914537231211030>.
- Cross, Ben. 2024. 'Taking Rulers' Interests Seriously: The Case for Realist Theories of Legitimacy'. *European Journal of Political Theory* 23 (2): 159–181. <https://doi.org/10.1177/14748851221114302>.
- Cudd, Ann, and Seena Eftekhari. 2025. 'Contractarianism'. In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Nodelman, Uri. <https://plato.stanford.edu/archives/fall2025/entries/contractarianism>.
- Daigle, Christine. 2006. 'Nietzsche: Virtue Ethics. . . Virtue Politics?' *The Journal of Nietzsche Studies* 32 (1): 1–21. <https://doi.org/10.1353/nie.2006.0012>.
- Dancy, Jonathan. 2006. *Ethics without Principles*. Reprinted. Clarendon Press.

- Delapp, Kevin M. n.d. 'Metaethics'. In *Internet Encyclopedia of Philosophy*. Accessed 19 July 2026. <https://iep.utm.edu/metaethi/>.
- Dennis, Matthew. 2019. 'Nietzschean Self-Cultivation: Connecting His Virtues to His Ethical Ideal'. *The Journal of Value Inquiry* 53 (1): 55–73. <https://doi.org/10.1007/s10790-018-9635-z>.
- Dewey, John. (1939) 2021. 'Creative Democracy— The Task Before Us'. In *America's Public Philosopher*, by John Dewey. Edited by Eric Thomas Weber. Columbia University Press. <https://doi.org/10.7312/dewe19894-011>.
- Dowding, Keith. 2020. 'The Relationship between Political Philosophy and Political Science'. *Australian Journal of Political Science* 55 (4): 432–444. <https://doi.org/10.1080/10361146.2020.1822776>.
- Driver, Julia. 2001. *Uneasy Virtue*. Cambridge Studies in Philosophy. Cambridge University Press. <https://doi.org/10.1017/CBO9780511498770>.
- Driver, Julia. 2003. 'The Conflation of Moral and Epistemic Virtue'. *Metaphilosophy* 34 (3): 367–383. <https://doi.org/10.1111/1467-9973.00279>.
- Dworkin, Ronald. 2011. *Justice for Hedgehogs*. 1. Harvard Univ. Press paperback ed. Belknap Press of Harvard University Press.
- Elkin, Stephen L. 2006. *Reconstructing the Commercial Republic: Constitutional Design after Madison*. University of Chicago Press.
- Erman, Eva. 2009. 'What Is Wrong with Agonistic Pluralism?: Reflections on Conflict in Democratic Theory'. *Philosophy & Social Criticism* 35 (9): 1039–1062. <https://doi.org/10.1177/0191453709343385>.
- Erman, Eva, and Niklas Möller. 2015. 'Why Political Realists Should Not Be Afraid of Moral Values'. *Journal of Philosophical Research* 40: 459–464. <https://doi.org/10.5840/jpr201511538>.
- Erman, Eva, and Niklas Möller. 2022. 'Distinctively Political Normativity in Political Theory'. *Philosophy Compass* 17 (6). <https://doi.org/10.1111/phc3.12835>.
- Erman, Eva, and Niklas Möller. 2023a. 'Does Normative Behaviourism Offer an Alternative Methodology in Political Theory?' *Political Studies Review* 21 (3): 454–461. <https://doi.org/10.1177/14789299231151803>.
- Erman, Eva, and Niklas Möller. 2023b. 'The Problem of Political Normativity Understood as Functional Normativity in Advance'. *Journal of Philosophical Research*, ahead of print. <https://doi.org/10.5840/jpr2023925212>.
- Erman, Eva, and Niklas Möller. 2023c. 'What Is Distinctive of Political Normativity? From Domain View to Role View'. *Philosophy & Social Criticism* 49 (3): 289–308. <https://doi.org/10.1177/01914537221131576>.
- Erman, Eva, and Niklas Möller. 2025. 'The Role of Moral Norms in Political Theory'. *Topoi* 44 (1): 27–38. <https://doi.org/10.1007/s11245-024-10066-3>.
- Farrelly, Colin. 2007. 'Justice in Ideal Theory: A Refutation'. *Political Studies* 55 (4): 844–864. <https://doi.org/10.1111/j.1467-9248.2007.00656.x>.

- Finlayson, Lorna. 2017. 'With Radicals like These, Who Needs Conservatives? Doom, Gloom, and Realism in Political Theory'. *European Journal of Political Theory* 16 (3): 264–282. <https://doi.org/10.1177/1474885114568815>.
- Floyd, Jonathan. 2011. 'From Historical Contextualism, to Mentalism, to Behaviourism'. In *Political Philosophy versus History?*, 1st edn, edited by Jonathan Floyd and Marc Stears. Cambridge University Press. <https://doi.org/10.1017/CBO9781139003698.003>.
- Floyd, Jonathan. 2017. *Is Political Philosophy Impossible?: Thoughts and Behaviour in Normative Political Theory*. 1st edn. Cambridge University Press. <https://doi.org/10.1017/9781316091081>.
- Floyd, Jonathan. 2020. 'Normative Behaviourism as a Solution to Four Problems in Realism and Non-Ideal Theory'. *Critical Review of International Social and Political Philosophy* 23 (2): 137–162. <https://doi.org/10.1080/13698230.2018.1501538>.
- Floyd, Jonathan. 2023a. 'Can Real Actions Justify Realist Principles? Normative Behaviourism as a Member of the Realist Family'. *Critical Review of International Social and Political Philosophy* 26 (3): 356–375. <https://doi.org/10.1080/13698230.2022.2120656>.
- Floyd, Jonathan. 2023b. 'Normative Behaviourism: A Reply'. *Political Studies Review* 21 (3): 490–500. <https://doi.org/10.1177/14789299231182092>.
- Foot, Philippa Ruth. 2001. *Natural Goodness*. Clarendon press.
- Foot, Philippa Ruth. 2002. *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford University Press. <https://doi.org/10.1093/0199252866.001.0001>.
- Foot, Philippa Ruth. 2004. 'Rationality and Goodness'. In *Modern Moral Philosophy*, 1st edn, edited by Anthony O'Hear. Cambridge University Press. <https://doi.org/10.1017/CBO9780511550836.002>.
- Fossen, Thomas. 2008. 'Agonistic Critiques of Liberalism: Perfection and Emancipation'. *Contemporary Political Theory* 7 (4): 376–394. <https://doi.org/10.1057/cpt.2008.15>.
- Foucault, Michel. 1978. *The History of Sexuality*. 1st American ed. With Gene Berry and Jeffrey Campbell Collection (Library of Congress). Pantheon Books.
- Franco, Paul. 2018. 'Becoming Who You Are: Nietzsche on Self-Creation'. *The Journal of Nietzsche Studies* 49 (1): 52–77. <https://doi.org/10.5325/jnietstud.49.1.0052>.
- Frankena, William K. 1973. *Ethics*. 2d ed. Prentice-Hall Foundations of Philosophy Series. Prentice-Hall.
- Frazer, Elizabeth. 2018. 'Feminism and Realism'. In *Politics Recovered*, edited by Matt Sleat. Columbia University Press. <https://doi.org/10.7312/sleat17528-015>.
- Freeden, Michael. 1996. *Ideologies and Political Theory: A Conceptual Approach*. 1. issued in paperback. Oxford University Press.
- Galston, William A. 2010. 'Realism in Political Theory'. *European Journal of Political Theory* 9 (4): 385–411. <https://doi.org/10.1177/1474885110374001>.
- Gert, Bernard, and Joshua Gert. 2025. 'The Definition of Morality'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Nodelman, Uri. <https://plato.stanford.edu/archives/spr2025/entries/morality-definition>.

- Geuss, Raymond. 2008. *Philosophy and Real Politics*. Princeton University Press.
- Geuss, Raymond. 2017. 'Realism, Wishful Thinking, Utopia'. In *Political Uses of Utopia*, edited by S. D. Chrostowska and James D. Ingram. Columbia University Press. <https://doi.org/10.7312/chro17958-013>.
- Glăveanu, Vlad P., and Ronald A. Beghetto. 2021. 'Creative Experience: A Non-Standard Definition of Creativity'. *Creativity Research Journal* 33 (2): 75–80. <https://doi.org/10.1080/10400419.2020.1827606>.
- Gledhill, James. 2012. 'Rawls and Realism': *Social Theory and Practice* 38 (1): 55–82. <https://doi.org/10.5840/soctheorpract20123813>.
- Goldie, Peter. 2000. *The Emotions: A Philosophical Exploration*. Clarendon Press.
- Gordon, John-Stewart. 2013. 'Modern Morality and Ancient Ethics'. In IEP. <https://iep.utm.edu/modern-morality-ancient-ethics/>.
- Grcic, Joseph. 2007. 'Hobbes and Rawls on Political Power'. *Etica & Politica / Ethics & Politics* 9 (2): 371–392.
- Habermas, Jurgen. 1995. 'Reconciliation Through the Public Use of Reason: Remarks on John Rawls's Political Liberalism'. *The Journal of Philosophy* 92 (3): 109. <https://doi.org/10.2307/2940842>.
- Haidt, Jonathan. 2001. 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment.' *Psychological Review* 108 (4): 814–834. <https://doi.org/10.1037/0033-295X.108.4.814>.
- Hall, Edward. 2020. *Value, Conflict, and Order: Berlin, Hampshire, Williams, and the Realist Revival in Political Theory*. The university of Chicago press.
- Hall, Edward, and Matt Sleat. 2017. 'Ethics, Morality and the Case for Realist Political Theory'. *Critical Review of International Social and Political Philosophy* 20 (3): 278–295. <https://doi.org/10.1080/13698230.2017.1293343>.
- Hatab, Lawrence J. 1987. 'Nietzsche, Nihilism and Meaning'. *The Personalist Forum* 3 (2): 91–111. <https://doi.org/10.5840/persforum1987322>.
- Hatab, Lawrence J. 2000. 'Nietzsche, Democracy, and Excellence: Politics as Jazz'. *International Studies in Philosophy* 32 (3): 39–50. <https://doi.org/10.5840/intstudphil200032339>.
- Hatab, Lawrence J. 2008. *Nietzsche's on the Genealogy of Morality: An Introduction*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511812002>.
- Hegel, Georg Wilhelm Friedrich. (1821) 1991. *Elements of the Philosophy of Right*. Edited by Allen William Wood. Translated by H. B. Nisbet. Cambridge Texts in the History of Political Thought. Cambridge university press.
- Helm, Bennett W. 2007. *Emotional Reason: Deliberation, Motivation, and the Nature of Value*. Digitally print. version. Cambridge Studies in Philosophy. Cambridge University Press.
- Herman, Barbara. 1993. *The Practice of Moral Judgment*. Harvard university press.
- Honig, Bonnie. (1993) 2023. *Political Theory and the Displacement of Politics*. 30th anniversary edition. Cornell University Press.

- Honig, Bonnie, and Alyson Cole. 2021. 'Ask A Political Scientist: A Discussion with Bonnie Honig on Agonism, Disruption, and Repair'. *Polity* 53 (2): 344–356. <https://doi.org/10.1086/712756>.
- Honig, Bonnie, and Marc Stears. 2011. 'The New Realism : From Modus Vivendi to Justice'. In *Political Philosophy versus History?*, 1st edn, edited by Jonathan Floyd and Marc Stears. Cambridge University Press. <https://doi.org/10.1017/CBO9781139003698.009>.
- Honig, Bonnie, and Marc Stears. 2014. 'James Tully's Agonistic Realism'. In *On Global Citizenship: James Tully in Dialogue*, 1st edn, by James Tully. Bloomsbury Publishing Plc. <https://doi.org/10.5040/9781472544827>.
- Hunt, Lester H. 1993. 'The Eternal Recurrence and Nietzsche's Ethic of Virtue.' *International Studies in Philosophy* 25 (2): 3–11. <https://doi.org/10.5840/intstudphil199325259>.
- Hunt, Lester H. (1991) 2005. *Nietzsche and the Origin of Virtue*. Taylor and Francis.
- Hurka, Thomas. 2001. *Virtue, Vice, and Value*. Oxford University Press, Incorporated.
- Hursthouse, Rosalind. 1999. *On Virtue Ethics*. Oxford University Press.
- Hursthouse, Rosalind, and Glen Pettigrove. 2023. 'Virtue Ethics'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman. <https://plato.stanford.edu/archives/fall2023/entries/ethics-virtue/>.
- Huxley, Aldous. (1931) 2006. *Brave New World*. Harper Perennial.
- Jubb, Robert. 2019. 'On What a Distinctively Political Normativity Is'. *Political Studies Review* 17 (4): 360–369. <https://doi.org/10.1177/1478929919832251>.
- Kant, Immanuel. (1785) 2008. *Groundwork for the Metaphysics of Morals*. Edited and translated by Allen W. Wood. With Marcia Baron, Shelly Kagan, J. B. Schneewind, and Allen W. Wood. Rethinking the Western Tradition. Yale University Press. <https://doi.org/10.12987/9780300128154>.
- Kapoor, Ilan. 2002. 'Deliberative Democracy or Agonistic Pluralism? The Relevance of the Habermas-Mouffe Debate for Third World Politics'. *Alternatives: Global, Local, Political* 27 (4): 459–487. <https://doi.org/10.1177/030437540202700403>.
- Kaufmann, Walter Arnold. 1974. *Nietzsche: Philosopher, Psychologist, Antichrist*. 4. ed., 20th pr. Princeton University Press.
- Korsgaard, Christine M. 1986. 'Kant's Formula of Humanity'. *Kant-Studien* 77 (1–4). <https://doi.org/10.1515/kant.1986.77.1-4.183>.
- Korsgaard, Christine M. 2008. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford University Press.
- Krastev, Ivan, and Mark Leonard. 2024. 'Wars and Elections: How European Leaders Can Maintain Public Support for Ukraine'. European Council on Foreign Relations, February 21. <https://ecfr.eu/publication/wars-and-elections-how-european-leaders-can-maintain-public-support-for-ukraine/>.
- Kreutz, Adrian. 2023. 'Realism and Metanormativity'. *Inquiry*, March 4, 1–29. <https://doi.org/10.1080/0020174X.2023.2185907>.

- Kupfer, Joseph. 2011. 'When the Badness of Vice Outweighs the Goodness of Virtue: A Moral Asymmetry'. *The Journal of Value Inquiry* 45 (4): 433–441. <https://doi.org/10.1007/s10790-011-9305-x>.
- Laclau, Ernesto, and Chantal Mouffe. (1985) 2001. *Hegemony and Socialist Strategy: Towards a Radical Democratic Politics*. 2nd ed. Verso.
- Landemore, Hélène. 2013. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton University Press. <https://www.jstor.org/stable/j.ctt1r2gf0>.
- Leader Maynard, Jonathan. 2022. 'Political Realism as Methods Not Metaethics'. *Ethical Theory and Moral Practice* 25 (3): 449–463. <https://doi.org/10.1007/s10677-021-10183-7>.
- Leader Maynard, Jonathan. 2024. 'The Logic of Idealization in Political Theory'. *American Journal of Political Science*, May 24, ajps.12869. <https://doi.org/10.1111/ajps.12869>.
- Leader Maynard, Jonathan. 2025. 'What Is Political Moralism?' *Topoi* 44 (1): 13–26. <https://doi.org/10.1007/s11245-024-10060-9>.
- Leader Maynard, Jonathan, and Alex Worsnip. 2018. 'Is There a Distinctively Political Normativity?' *Ethics* 128 (4): 756–787. <https://doi.org/10.1086/697449>.
- LeBar, Mark. 2009. 'Virtue Ethics and Deontic Constraints'. *Ethics* 119 (4): 642–671. <https://doi.org/10.1086/603649>.
- LeBar, Mark. 2013. *The Value of Living Well*. Oxford University Press.
- Leiter, Brian. 1997. 'Nietzsche and the Morality Critics'. *Ethics* 107 (2): 250–285. <https://doi.org/10.1086/233720>.
- Leiter, Brian. 2024. 'Nietzsche's Moral and Political Philosophy'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman. <https://plato.stanford.edu/archives/fall2024/entries/nietzsche-moral-political/>.
- Lévinas, Emmanuel. 1996. *Emmanuel Levinas: Basic Philosophical Writings*. Edited by Adriaan Theodoor Peperzak, Simon Critchley, and Robert Bernasconi. Studies in Continental Thought. Indiana University Press.
- Mac Cumhaill, Clare, and Rachael Wiseman. 2022. *Metaphysical Animals: How Four Women Brought Philosophy Back to Life*. First American Edition. Doubleday.
- MacIntyre, Alasdair C. (1981) 2007. *After Virtue: A Study in Moral Theory*. 3rd ed. University of Notre Dame Press.
- Magnus, Bernd. 1980. 'Aristotle and Nietzsche: "Megalopsychia" and "Uebermensch"'. In *The Greeks and the Good Life*, by David J. Depew. California State Fullerton Press.
- Markell, Patchen. 2003. *Bound by Recognition*. Princeton University Press.
- Matthew, D. C. 2019. 'Rawls's Ideal Theory: A Clarification and Defense'. *Res Publica* 25 (4): 553–570. <https://doi.org/10.1007/s11158-018-09412-1>.
- McDowell, John. 1979. 'Virtue and Reason'. *Monist* 62 (3): 331–350. <https://doi.org/10.5840/monist197962319>.
- McQueen, Alison. 2017. 'Political Realism and the Realist "Tradition"'. *Critical Review of International Social and Political Philosophy* 20 (3): 296–313. <https://doi.org/10.1080/13698230.2017.1293914>.

- Meijer, Michiel. 2022. 'Articulating Better, Being Better: Ethical Emancipation and the Sources of Motivation'. *Ethical Theory and Moral Practice* 25 (1): 107–122.
<https://doi.org/10.1007/s10677-021-10233-0>.
- Metz, Thaddeus. 2023. 'The Meaning of Life'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Nodelman, Uri.
<https://plato.stanford.edu/archives/fall2023/entries/life-meaning/>.
- Miyasaki, Donovan. 2022a. *Nietzsche's Immoralism: Politics as First Philosophy*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-11359-8>.
- Miyasaki, Donovan. 2022b. *Politics after Morality: Toward a Nietzschean Left*. Palgrave Macmillan.
- Montmarquet, James A. 1993. *Epistemic Virtue and Doxastic Responsibility*. Studies in Epistemology and Cognitive Theory. Rowman & Littlefield.
- Moore, G. E. (1903) 2002. *Principia Ethica: With the Preface to the Second Edition and Other Papers*. Edited by Thomas Baldwin. Repr., Transf. to digital print. Cambridge Univ. Press.
- Mouffe, Chantal. 1999. 'Deliberative Democracy or Agonistic Pluralism?' *Social Research* 66 (3): 745–758.
- Mouffe, Chantal. 2000. *The Democratic Paradox*. Verso.
- Mouffe, Chantal. 2005. 'The Limits of John Rawls's Pluralism'. *Politics, Philosophy & Economics* 4 (2): 221–231. <https://doi.org/10.1177/1470594x05052539>.
- Mouffe, Chantal. 2013. *Agonistics: Thinking the World Politically*. [CA & US version]. With Elke Wagner. Verso.
- Müller, Jozef. 2015. 'Aristotle on Vice'. *British Journal for the History of Philosophy* 23 (3): 459–477.
<https://doi.org/10.1080/09608788.2015.1022855>.
- Murdoch, Iris. (1970) 2014. *The Sovereignty of Good*. With Mary Midgley. Routledge Great Minds Series. Routledge.
- Nehamas, Alexander. 2001. 'How One Becomes What One Is'. In *Nietzsche*, edited by John Richardson and Brian Leiter. Oxford Readings in Philosophy. Oxford University Press.
- Nietzsche, Friedrich. 1888. *Nietzsche Contra Wagner Out of the Files of a Psychologist*.
<http://www.thenietzschechannel.com/works-pub/ncw/ncw.htm>.
- Nietzsche, Friedrich. (1888) 1923. *The Antichrist*. Translated by H. L. Mencken. ALFRED A. KNOPF / The Project Gutenberg EBook. <https://www.gutenberg.org/files/19322/19322-h/19322-h.htm>.
- Nietzsche, Friedrich. (1882) 1974. *The Gay Science: With a Prelude in Rhymes and an Appendix of Songs*. Translated by Walter Arnold Kaufmann. Vintage Books.
- Nietzsche, Friedrich. 1988. *Nachgelassene Fragmente 1880–1882 (Band 9)*. Edited by Giorgio Colli and Mazzino Montinari. De Gruyter. <https://doi.org/10.1515/9783112419021>.
- Nietzsche, Friedrich. (1881) 1997. *Daybreak: Thoughts on the Prejudices of Morality*. Edited by Maudemarie Clark and Brian Leiter. Cambridge Texts in the History of Philosophy. Cambridge university press.

- Nietzsche, Friedrich. (1886) 1998. *Beyond Good and Evil: Prelude to a Philosophy of the Future*. Translated by Marion Faber. Oxford World's Classics. Oxford University Press.
- Nietzsche, Friedrich. (1883) 2006. *Thus Spoke Zarathustra: A Book for All and None*. Edited by Adrian Del Caro and Robert B. Pippin. Translated by Adrian Del Caro. Cambridge Texts in the History of Philosophy. Reprint with corrections. Cambridge Univ. Press.
- Nietzsche, Friedrich. (1887) 2007. *On the Genealogy of Morality*. Edited by Keith Ansell-Pearson. Translated by Carol Diethe. Cambridge Texts in the History of Political Thought. Revised ed. Cambridge university press.
- Norval, Aletta J. 2007. *Aversive Democracy: Inheritance and Originality in the Democratic Tradition*. Cambridge University Press.
- Nussbaum, Martha C. 2002. 'Rawls and Feminism'. In *The Cambridge Companion to Rawls*, 1st edn, edited by Samuel Freeman. Cambridge University Press.
<https://doi.org/10.1017/CCOL0521651670.015>.
- Nussbaum, Martha C. 2011. *Creating Capabilities: The Human Development Approach*. Belknap Press of Harvard University Press.
- Nye, Sebastian. 2015. 'Real Politics and Metaethical Baggage'. *Ethical Theory and Moral Practice* 18 (5): 1083–1100. <https://doi.org/10.1007/s10677-015-9590-8>.
- Oakeshott, Michael. 1962. *Rationalism in Politics, and Other Essays*. Basic Books.
- Okin, Susan Moller. 1989. *Justice, Gender and the Family*. Basic books.
- O'Neill, Onora. 1996. *Towards Justice and Virtue: A Constructive Account of Practical Reasoning*. Transferred to digital print. Cambridge University Press.
- O'Neill, Onora. 2003. *A Question of Trust*. The BBC Reith Lectures. Cambridge University Press.
- Ostrom, Elinor. 2000. 'Reformulating the Commons'. *Swiss Political Science Review* 6 (1): 29–52.
<https://doi.org/10.1002/j.1662-6370.2000.tb00285.x>.
- Owen, David. 1996. 'Agonal Thought: Reading Nietzsche as Political Thinker'. *Angelaki* 1 (3): 119–128. <https://doi.org/10.1080/09697259608571901>.
- Owen, David. 2002. 'Equality, Democracy, and Self-Respect: Reflections on Nietzsche's Agonal Perfectionism'. *The Journal of Nietzsche Studies* 24 (1): 113–131.
<https://doi.org/10.1353/nie.2002.0014>.
- Owen, David. 2018. 'Realism in Ethics and Politics: Bernard Williams, Political Theory, and the Critique of Morality'. In *Politics Recovered*, edited by Matt Sleat. Columbia University Press. <https://doi.org/10.7312/sleat7528-005>.
- Perez, Nahshon. 2020. 'What Are Data Good for Anyway?: A Typology of Usages of Data in Contemporary Political Theory'. *Social Theory and Practice* 46 (2): 339–364.
<https://doi.org/10.5840/soctheorpract20204188>.
- Perez, Nahshon. 2024. 'Empirical Political Theory: A Template for a Research Design and a Qualified Defense'. *Political Studies Review*, February 26, 14789299241231782.
<https://doi.org/10.1177/14789299241231782>.
- Philp, Mark. 2010. 'What Is to Be Done? Political Theory and Political Realism'. *European Journal of Political Theory* 9 (4): 466–484. <https://doi.org/10.1177/1474885110374010>.

- Piirimäe, Eva. 2020. '3 Human Rights and Their Realisation in the World: Herder's Debate with Kant'. In *Passions, Politics and the Limits of Society*, edited by Heikki Haara, Koen Stapelbroek, and Mikko Immanen. De Gruyter. <https://doi.org/10.1515/9783110679861-004>.
- Power, Michael. 1999. *The Audit Society: Rituals of Verification*. Reprint. Oxford University Press.
- Prinz, Janosch, and Enzo Rossi. 2017. 'Political Realism as Ideology Critique'. *Critical Review of International Social and Political Philosophy* 20 (3): 348–365. <https://doi.org/10.1080/13698230.2017.1293908>.
- Przeworski, Adam, Susan Carol Stokes, and Bernard Manin, eds. 1999. *Democracy, Accountability, and Representation*. Cambridge Studies in the Theory of Democracy 2. Cambridge University Press. <https://doi.org/10.1017/CBO9781139175104>.
- Quong, Jonathan. 2011. *Liberalism without Perfection*. First edition. Oxford University Press.
- Raekstad, Paul. 2021. 'The Model of the Legislator: Political Theory, Policy, and Realist Utopianism'. *Contemporary Political Theory* 20 (4): 727–748. <https://doi.org/10.1057/s41296-021-00469-x>.
- Rawls, John. 1974. 'The Independence of Moral Theory'. *Proceedings and Addresses of the American Philosophical Association* 48: 5. <https://doi.org/10.2307/3129858>.
- Rawls, John. 1985. 'Justice as Fairness: Political Not Metaphysical'. *Philosophy and Public Affairs* 14 (3): 223–251.
- Rawls, John. 1996. *Political Liberalism*. Paperback ed. The John Dewey Essays in Philosophy 4. Columbia Univ. Press.
- Rawls, John. (1971) 1999. *A Theory of Justice*. Rev. ed. Belknap Press of Harvard University Press.
- Rawls, John. (1975) 2001. 'A Kantian Conception of Equality'. In *Collected Papers*, edited by Samuel Freeman. Harvard University Press. <http://www.jstor.org/stable/10.2307/j.ctv1cbn3j4>.
- Rawls, John. 2003. *Justice as Fairness: A Restatement*. 3. printing. Edited by Erin Kelly. Belknap Press of Harvard University Press.
- Robson, Gregory. 2024. 'The Varieties of Moral Vice: An Aristotelian Approach'. *Erkenntnis* 89 (5): 1993–2012. <https://doi.org/10.1007/s10670-022-00614-x>.
- Rogers, Tristan J., and Daniel C. Russell. 2025. 'Virtue Ethics and Political Philosophy'. In *The Routledge Companion to Social and Political Philosophy*, Second edition, edited by Gerald F. Gaus, Fred D'Agostino, and Ryan Muldoon. Routledge Philosophy Companions. Routledge. <https://doi.org/10.4324/9781003411598>.
- Rønnow-Rasmussen, Toni. 2015. *Intrinsic and Extrinsic Value*. Edited by Iwao Hirose and Jonas Olson. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199959303.013.0003>.
- Rosati, Connie. 2016. 'Moral Motivation'. In *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/moral-motivation/>.
- Rossi, Enzo. 2019. 'Being Realistic and Demanding the Impossible'. *Constellations* 26 (4): 638–652. <https://doi.org/10.1111/1467-8675.12446>.

- Rossi, Enzo. 2023. 'Fact-Centric Political Theory, Three Ways: Normative Behaviourism, Grounded Normative Theory, and Radical Realism'. *Political Studies Review* 21 (3): 483–489. <https://doi.org/10.1177/14789299231157625>.
- Rossi, Enzo. 2025. 'What Can Epistemic Normativity Tell Us About Politics? Ideology, Power, and the Epistemology of Radical Realism'. *Topoi* 44 (1): 77–88. <https://doi.org/10.1007/s11245-024-10142-8>.
- Rossi, Enzo, and Carlo Argenton. 2021. 'Property, Legitimacy, Ideology: A Reality Check'. *The Journal of Politics* 83 (3): 1046–1059. <https://doi.org/10.1086/710781>.
- Rossi, Enzo, and Matt Sleat. 2014. 'Realism in Normative Political Theory: Realism in Normative Political Theory'. *Philosophy Compass* 9 (10): 689–701. <https://doi.org/10.1111/phc3.12148>.
- Rowe, M. W. 1991. 'Goethe and Wittgenstein'. *Philosophy* 66 (257): 283–303. <https://doi.org/10.1017/S0031819100064901>.
- Sandel, Michael J. (1982) 1998. *Liberalism and the Limits of Justice*. 2nd ed. Cambridge University Press.
- Saunders, Cicely. 2000. 'The Evolution of Palliative Care'. *Patient Education and Counseling* 41 (1): 7–13. [https://doi.org/10.1016/S0738-3991\(00\)00110-5](https://doi.org/10.1016/S0738-3991(00)00110-5).
- Sayre-McCord, Geoffrey. 2023. 'Metaethics'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman. <https://plato.stanford.edu/archives/spr2023/entries/metaethics>.
- Scanlon, Thomas. 2000. *What We Owe to Each Other*. First Harvard University Press paperback edition. The Belknap Press of Harvard University Press.
- Schneewind, J. B. 1998. *The Invention of Autonomy: A History of Modern Moral Philosophy*. Cambridge university press.
- Schoolman, Morton. (1993) 2002. 'Series Editor's Introduction'. In *The Augustinian Imperative: A Reflection on the Politics of Morality*, by William E. Connolly. Modernity and Political Thought. New edition. Rowman & Littlefield.
- Schwarzenbach, Sibyl A. 2009. *On Civic Friendship: Including Women in the State*. Columbia University Press.
- Scott, Joan Wallach. 2010. *The Politics of the Veil*. 8. printing and 1. paperback printing. The Public Square Book Series. Princeton Univ. Press.
- Selg, Peeter. 2012. 'Justice and Liberal Strategy: Towards a Radical Democratic Reading of Rawls'. *Social Theory and Practice* 38 (1): 83–114. <https://doi.org/10.5840/soctheorpract20123814>.
- Sen, Amartya. 1993. 'Capability and Well-Being'. In *The Quality of Life*, edited by Martha Nussbaum and Amartya Sen. Oxford University Press. <https://doi.org/10.1093/0198287976.003.0003>.
- Shklar, Judith. 1984. *Ordinary Vices*. The Belknap Press of Harvard University Press.
- Silvia, Paul J. 2008. 'Interest—The Curious Emotion'. *Current Directions in Psychological Science* 17 (1): 57–60. <https://doi.org/10.1111/j.1467-8721.2008.00548.x>.

- Singer, Peter. (1975) 2009. *Animal Liberation: The Definitive Classic of the Animal Movement*. 1st Harper Perennial ed. Ecco Book/Harper Perennial.
- Sleat, Matt. 2010. 'Bernard Williams and the Possibility of a Realist Political Theory'. *European Journal of Political Theory* 9 (4): 485–503. <https://doi.org/10.1177/1474885110382689>.
- Sleat, Matt. 2013. *Liberal Realism: A Realist Theory of Liberal Politics*. Manchester Univ. Press.
- Sleat, Matt. 2014. 'Legitimacy in Realist Thought: Between Moralism and Realpolitik'. *Political Theory* 42 (3): 314–337. <https://doi.org/10.1177/0090591714522250>.
- Sleat, Matt, ed. 2018. *Politics Recovered: Realist Thought in Theory and Practice*. Columbia University Press.
- Sleat, Matt. 2022. 'Realism and Political Normativity'. *Ethical Theory and Moral Practice* 25 (3): 465–478. <https://doi.org/10.1007/s10677-021-10239-8>.
- Sleat, Matt. 2024. 'Against Realist Ideology Critique'. *Social Philosophy and Policy* 41 (1): 139–157. <https://doi.org/10.1017/s0265052524000360>.
- Slote, Michael. 2001. 'Agent-Based Virtue Ethics'. In *Morals from Motives*, 1st edn, by Michael Slote. Oxford University Press. <https://doi.org/10.1093/0195138376.003.0001>.
- Slote, Michael. 2007. *The Ethics of Care and Empathy*. Routledge.
- Slote, Michael. 2010. *Moral Sentimentalism*. First issued as an Oxford University Press paperback. Oxford University Press.
- Slote, Michael. 2013. *From Enlightenment to Receptivity: Rethinking Our Values*. Oxford university press.
- Slote, Michael. 2023. 'Agent-Based Virtue Ethics'. In *Philosophical Essays East and West*, by Michael Slote. Palgrave Studies in Comparative East-West Philosophy. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-39955-8_6.
- Smiley, Marion. 2004. 'Gender, Democratic Citizenship v. Patriarchy: A Feminist Perspective on Rawls'. *Fordham Law Review* 5: 1599–1627.
- Sorgiovanni, Ben. 2020. 'Wittgenstein on the Constitutive Uncertainty of the Mental'. *Nordic Wittgenstein Review* 9 (September). <https://doi.org/10.15845/nwr.v0i0.3562>.
- Southwood, Nicholas. 2009. 'Moral Contractualism'. *Philosophy Compass* 4 (6): 926–937. <https://doi.org/10.1111/j.1747-9991.2009.00256.x>.
- Srinivasan, Amia. 2019. 'VII — Genealogy, Epistemology and Worldmaking'. *Proceedings of the Aristotelian Society* 119 (2): 127–156. <https://doi.org/10.1093/arisoc/aoz009>.
- Stahl, Titus. 2013. 'What Is Immanent Critique?' SSRN Electronic Journal, ahead of print. <https://doi.org/10.2139/ssrn.2357957>.
- Stahl, Titus. 2017. 'The Metaethics of Critical Theories'. In *The Palgrave Handbook of Critical Theory*, by Michael J. Thompson. Palgrave Handbooks. Palgrave Macmillan.
- Stahl, Titus. 2022a. 'Immanent Critique and Particular Moral Experience'. *Critical Horizons* 23 (1): 1–21. <https://doi.org/10.1080/14409917.2017.1376939>.
- Stahl, Titus. 2022b. *Immanent Critique*. Translated by John-Baptiste Oduor. Essex Studies in Contemporary Critical Theory. Rowman & Littlefield.

- Stemplowska, Zofia. 2008. 'What's Ideal About Ideal Theory?:' *Social Theory and Practice* 34 (3): 319–340. <https://doi.org/10.5840/soctheorpract200834320>.
- Swanton, Christine. 2003. *Virtue Ethics: A Pluralistic View*. Reprinted. Oxford Univ. Press.
- Swanton, Christine. 2005. 'Nietzschean Virtue Ethics'. In *Virtue Ethics, Old and New*, edited by Stephen M. Gardiner. Cornell University Press. <https://doi.org/10.7591/9781501724275>.
- Swanton, Christine. 2015. *The Virtue Ethics of Hume and Nietzsche*. 1st edn. Wiley. <https://doi.org/10.1002/9781118939369>.
- Swanton, Christine. 2021. *Target-Centred Virtue Ethics*. First edition. Oxford University Press.
- Tangney, J. P., J. Stuewig, and D. J. Mashek. 2007. 'Moral Emotions and Moral Behavior'. *Annual Review of Psychology* 58: 345–372.
- Thomas, Alan. 2017. 'Rawls and Political Realism: Realistic Utopianism or Judgement in Bad Faith?' *European Journal of Political Theory* 16 (3): 304–324. <https://doi.org/10.1177/1474885115578970>.
- Toomela, Aaro. 2016. 'What Are Higher Psychological Functions?' *Integrative Psychological and Behavioral Science* 50 (1): 91–121. <https://doi.org/10.1007/s12124-015-9328-0>.
- Tully, James. 2008a. *Public Philosophy in a New Key. Volume 1: Democracy and Civic Freedom*. 1st edn. Cambridge University Press. <https://doi.org/10.1017/CBO9780511790737>.
- Tully, James. 2008b. *Public Philosophy in a New Key. Volume 2, Imperialism and Civic Freedom*. 1st edn. Cambridge University Press. <https://doi.org/10.1017/CBO9780511790744>.
- Turri, John, Mark Alfano, and John Greco. 2021. 'Virtue Epistemology'. In *The Stanford Encyclopedia of Philosophy*, 2021 Winter, edited by Edward N. Zalta. <https://plato.stanford.edu/archives/win2021/entries/epistemology-virtue>.
- Vallier, Kevin. 2017. 'Three Concepts of Political Stability: An Agent-Based Model'. *Social Philosophy and Policy* 34 (1): 232–259. <https://doi.org/10.1017/S0265052517000115>.
- Van Zyl, Liezl. 2009. 'Agent-Based Virtue Ethics and the Problem of Action Guidance'. *Journal of Moral Philosophy* 6 (1): 50–69. <https://doi.org/10.1163/174552409X365928>.
- Van Zyl, Liezl. 2011. 'Rightness and Goodness in Agent-Based Virtue Ethics'. *Journal of Philosophical Research* 36: 103–114. https://doi.org/10.5840/jpr_2011_3.
- Väyrynen, Pekka. 2011. 'Moral Particularism'. In *The Bloomsbury Companion to Ethics*, by Christian B. Miller. Bloomsbury Companions. Bloomsbury. <https://philarchive.org/archive/VYRMP>.
- Vittersø, Joar. 2025. 'Eudaimonia'. In *Humanistic Wellbeing*, by Joar Vittersø. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-69292-5_6.
- Wenman, Mark. 2013. *Agonistic Democracy: Constituent Power in the Era of Globalisation*. 1st edn. Cambridge University Press. <https://doi.org/10.1017/CBO9780511777158>.
- Weston, Michael, ed. 1994. *Kierkegaard and Modern Continental Philosophy: An Introduction*. Routledge.
- Westphal, Manon. 2023. 'Against the Status Quo: The Social as a Resource of Critique in Realist Political Theory'. *Critical Review of International Social and Political Philosophy* 26 (3): 418–436. <https://doi.org/10.1080/13698230.2022.2120660>.

- Wiens, David. 2015. 'Against Ideal Guidance'. *The Journal of Politics* 77 (2): 433–446.
<https://doi.org/10.1086/679495>.
- Williams, Bernard. 1985. *Ethics and the Limits of Philosophy*. Transferred to digital print.
 Routledge.
- Williams, Bernard. 2002. *Truth & Truthfulness: An Essay in Genealogy*. Princeton Univ. Press.
- Williams, Bernard. 2005. In *the Beginning Was the Deed: Realism and Moralism in Political Argument*. Edited by Geoffrey Hawthorn. Princeton University Press.
- Wilson, Alan T. 2017. 'Avoiding the Conflation of Moral and Intellectual Virtues'. *Ethical Theory and Moral Practice* 20 (5): 1037–1050. <https://doi.org/10.1007/s10677-017-9843-9>.
- Wittgenstein, Ludwig. 2010. *Philosophische Untersuchungen: = Philosophical investigations*. Rev. 4. ed., 3. [Dr.]. Translated by G. E. M. Anscombe, P. M. S. Hacker, and Joachim Schulte. Wiley-Blackwell.
- Young, Iris Marion. 1990. *Justice and the Politics of Difference*. Princeton University Press.
- Zagzebski, Linda. 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge University Press.
- Zagzebski, Linda. 2005. 'Virtue Epistemology'. In *Routledge Encyclopedia of Philosophy*, 1st edn. Routledge. <https://doi.org/10.4324/9780415249126-P057-2>.
- Zembylas, Michalinos. 2021. 'Refusal as Affective and Pedagogical Practice in Higher Education Decolonization: A Modest Proposal'. *Teaching in Higher Education* 26 (7–8): 953–968.
<https://doi.org/10.1080/13562517.2021.1900816>.
- Zimmerman, Michael, and Ben Bradley. 2025. 'Intrinsic vs. Extrinsic Value'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta.
<https://plato.stanford.edu/archives/sum2025/entries/value-intrinsic-extrinsic/>.