

Reasons for Reasons

A Theory of Reasons for Concept Use

Habilitationsschrift

zur Erlangung der Venia docendi
der Philosophisch-historischen
Fakultät der Universität Bern

vorgelegt von

MATTHIEU QUELOZ

Doktor der Philosophie (Dr. phil)
von Saint-Brais, JU

2023

Abstract

Philosophy strives to give us a firmer hold on our concepts. But what about their hold on us? Why should we place ourselves under the sway of a particular definition and grant it the authority to shape our thought and conduct? A different conceptualization would carry different implications. What makes one concept better than another?

This book develops a framework for concept appraisal. Its guiding idea is that to question the authority of concepts is to demand *reasons for reasons*: second-order reasons to use certain concepts and heed the first-order reasons they advert to. Such reasons for concept use promise to shore up—or undermine—the reasons for action and belief that guide our deliberations. They promise to tell us which concepts to adopt, adhere to, or abandon.

The book advances two main claims. The first is that reasons for concept use are to be found in our conceptual needs: the needs for certain concepts we have by virtue of our concerns. The second is that sometimes, concepts that conflict, or exhibit other theoretical vices such as vagueness or superficiality, are just what we need.

By considering not what concepts are definitively best, but what concepts we now need, we can reconcile ourselves to the contingency of our concepts, identify the proper place of efforts to tidy up thought, and adjudicate between competing conceptions of contested notions like liberty or free will. A needs-based approach empowers us to distinguish helpful clarification from hamstringing tidy-mindedness and authoritative definition from conceptual gerrymandering.

Keywords: second-order reasons, conceptual ethics, thick concepts, politics, authority, contingency, knowledge, theoretical virtues, conflicts of value, pluralism, respect, liberty, free will, moral luck, action theory, Bernard Williams, Ronald Dworkin, Susan Wolf, Iris Murdoch.

Contents

Introduction: Appraising Concepts	i
The Authority Question	1
1.1 Dworkin's Challenge	1
1.2 The Power of Conceptual Architectures	4
1.3 Expressing the Authority Question	11
1.4 When the Authority Question Arises	15
1.5 The Autoethnographic Stance	21
1.6 Conflating Engaged and Disengaged Use	27
1.7 Concepts for Conceptual Ethics	32
Confidence, Reflection, and Knowledge	41
2.1 Confidence in Concepts	41
2.2 Metaconceptual Reflection	44
2.3 Knowledge under Concepts	55
2.4 Metaconceptual Knowledge	59
Escaping the Trilemma	63
3.1 Generalized Foundationalism	63
3.2 Indiscriminate Ironism	68
3.3 Reasons for Us: Non-Foundationalism	72
3.4 Undiscriminating Holism	77
3.5 The Kaleidoscopic Picture	82
3.6 Leveraging Local Needs	85
Tidy-Mindedness	96
4.1 Theoretical Vices in Concepts	98
4.2 Superficial Concepts	100
4.3 Conceptual Tensions	104
4.4 Authority through Theoretical Virtue	117
4.5 Inheriting Authority from Theories	121
Concepts and Concerns	134
5.1 The Dworkin–Williams Debate	134
5.2 Lessons from the Dworkin–Williams Debate	145
5.3 The Practical Virtues of Theoretical Vices	149
5.4 The Limits of Concerns: Four Problems	157

Tailoring Concepts to Needs	162
6.1 Conceptual Needs	162
6.2 Needfulness Conditions	166
6.3 The Expressive Character of Concepts	174
6.4 Need Matrices	181
6.5 Needs-Based Concept Appraisal	186
6.6 Four Problems Solved	192
 Reasons for Concept Use	 200
7.1 Reasons in vs. Reasons for Concept Use	200
7.2 Concern-Independent Reasons in Concept Use	205
7.3 Instrumentality Without Instrumental Mentality	208
7.4 From Concerns to Reasons in Concept Use	212
7.5 Answering the Authority Question	219
7.6 Wrong Kinds of Reasons	232
7.7 Conceptual Ethics vs. Conceptual Morality	239
7.8 Conceptual Good-for-Nothings	240
 The Essential Superficiality of the Voluntary	 245
8.1 A Questionable Concept	245
8.2 Making Sense and Knowing What to Expect	251
8.3 Fairness and Freedom	259
8.4 Knowledge and Coercion	262
8.5 When Concerns Distort Concepts	266
8.6 Deepening the Concept of the Voluntary	269
8.7 The First Problem of Free Will	272
8.8 The Second Problem of Free Will	278
 The Politics of Conflicting Concerns	 283
9.1 Three Demands of Political Disagreement	283
9.2 The Dworkin–Williams Debate Continued	286
9.3 Enabling Political Contestation	291
9.4 Respect Across the Aisle	295
9.5 Placing the Demand for Theoretical Virtues	301
 Conclusion	 317
 Bibliography	 321

Acknowledgements

Some of the ideas in this book were first tested in the form of journal articles, and I am grateful to the publishers for permission to draw on excerpts from these articles here. These are: ‘Function-Based Conceptual Engineering and the Authority Problem’, *Mind*, 131, no. 524 (2022), 1247–1278, by permission of Oxford University Press. Copyright © 2022. ‘The Essential Superficiality of the Voluntary and the Moralization of Psychology’, *Philosophical Studies*, 179, no. 5 (2022), 1591–1620, by permission of Springer Nature. Copyright © 2021. ‘Nietzsche’s Conceptual Ethics’, *Inquiry*, Published Online (2023), 1–31, by permission of Taylor and Francis. Copyright © 2023.

Parts of the book were also presented at the Universities of Oxford, Bern, Southampton, Lund, Warwick, Bologna, Cambridge, and Hong Kong as well as at the 9th International Lauener Symposium on Analytical Philosophy. I am grateful to the discussants on all these occasions. [Further acknowledgements to come.]

It is one thing to justify a thought on the basis of other thoughts—something else to justify thinking.

Ludwig Wittgenstein, *The Big Typescript*, 180

The responsible introduction or extension of terms, whether in philosophy or science, reflects a conceptual need.

Justus Buchler, *Nature and Judgment*, 108

Introduction: Appraising Concepts

While much philosophy strives to give us a firmer hold on our concepts, we sometimes also find ourselves questioning their hold on us: why should we place ourselves under their sway and grant them the authority to shape our thought and conduct?¹ The concepts we use render us sensitive to the reasons that guide and flow from their application. But what reasons do we have to heed those reasons in the first place? If our thoughts were cast in different terms, they would advert to different reasons, carry different implications, and set us on different trajectories. Do our conceptualizations—the ways of thinking and valuing in virtue of which we possess the concepts we do—merit the confidence with which we deploy them? What makes one conceptualization better than another?

The question matters, because not every issue, in philosophy or elsewhere, consists simply in figuring out what is true or what is justified given the way we conceptualize things. Many issues are, at least in part, about how to conceptualize things—how to carve things up, how to characterize them, and what significance to attach to them. People can form perfectly true and justified judgements and nonetheless attract criticism for the very

¹ The question is significantly different from the Kantian, semantic concern to understand how concept application can be liable to assessments of correctness that Robert Brandom, drawing a *prima facie* similar chiastic contrast, sees as replacing the Cartesian, epistemological concern with whether our ideas are clear and distinct: ‘For Kant the question is … how to understand [concepts’] grip on us: the conditions of the intelligibility of our being bound by conceptual norms’ (2009, 33); see also Brandom (1994, 9–11; 2000, 80; 2002a, 22; 2019b, 9). My question is not *how any* concepts can bind us, but *why these* concepts rather than others should be allowed to.

terms in which they think.

Questioning the terms in which we think goes to the root of our thinking processes, for while the judgements we form might be criticized as false or unwarranted, these criticisms still take for granted the conceptual framework within which those judgements are articulated. By contrast, appraising the concepts we possess—by which I mean our concepts *as we actually understand and use them*, which is to say our *conceptualizations*—goes one step further, asking whether things might not go better if we reasoned along different lines that put alternative sets of judgements and patterns of justification within our reach. We might find mistakes in the way a computer executes the rules it operates by, or malfunctions in the way a smartphone runs its apps, but the deeper critique is the one that criticizes the very rules the computer attempts to apply, or the very apps the smartphone seeks to run. Analogously, our conceptualizations can fruitfully be regarded as pieces of mindware, encoding certain ways of thinking by scripting appropriate patterns of thought.² To ask which concepts we should use is to ask what mindware society should run on.

It is this demand for reasons to cast our thought in certain terms rather than others that forms my topic in this book. The concepts we use determine what we recognize as a reason for what, but these tend to be reasons for belief and reasons for action. What about *reasons for concept use*? Are there reasons for us to employ some concepts in preference to others? These would have to be *reasons for reasons*, which is to say second-order reasons to use certain concepts and be responsive to the first-order reasons they advert to.

Since our conceptual repertoire is a motley mixture of thinking techniques that vary widely in how they work and what purposes they serve, some concepts are more liable than others to invite demands for reasons to prefer them over alternatives. Among the concepts that are especially liable to do so are the concepts I shall focus on in this book, which are the concepts that unite the following three features: they are *world-guided*, meaning that their application is closely guided by how the world is rather than by our will; they are *action-guiding*, meaning that their applicability typically gives us reasons for action; and they are culturally *local* rather than universal, meaning that they compete with

² I take the term ‘mindware’ from Clark (2013), though it also figures prominently in Nisbett (2015). A closely related metaphor is J. L. Balkin’s (1998) notion of ‘cultural software’, which in turn echoes Clifford Geertz’s notion of ‘cultural templates’ (1973a, 217–18).

alternative concepts, real and imagined, for a role in shaping our lives.

Concepts combining these gradable features to any considerable degree are sometimes called ‘thick’ normative concepts.³ The concepts *blasphemy*, *heroism*, or *treason* are examples; so, to take more recent additions to the conceptual repertoire, are *mansplaining*, *manspreading*, or *himpathy*. These concepts are ‘thick’ because they are thickly descriptive—they have a higher descriptive content than thinner ones like *rational*, *good*, or *right* (if all one is told about x is that x was *disloyal* or *cowardly*, one still has a far more determinate idea of what x is and what happened than if told only that x was *bad* or *wrong*). But these concepts are also ‘normative’ in that they do more than describe or pick out things. As Bernard Williams notes, ‘what your repertoire of thick concepts is reveals your own or your society’s ethical attitude’ (1995k, 237), because to think in terms of concepts like *kitsch*, *sacrilegious*, *chaste*, or *unpatriotic* is not just to be sensitive to the presence of things that fall under these concepts, but to cast these things in a certain evaluative light. That in turn makes a difference to what attitude one has reason to adopt towards them, and, ultimately, to what one has reason to do.

This action-guiding import of certain concepts is something we overlook if we appraise concepts merely according to whether they match up with an antecedently articulated world.⁴ On Theodore Sider’s account, for example, if a community has true and warranted beliefs but nonetheless has ‘the wrong concepts’, this must be because these concepts do not match the world’s ‘structure’ (2011, 2). But Sider primarily has the concepts of fundamental physics in mind. One’s willingness to extend this approach to thick normative concepts will depend on whether one regards the ‘structure’ corresponding to thick normative concepts as sufficiently independent of those concepts

³ The notion of a thick concept and the world-guided/action-guiding terminology is associated notably with Bernard Williams (2011, 143–45), who is indebted in this connection to Wittgensteinian ideas developed by Philippa Foot and Iris Murdoch in a seminar in the 1950s (Williams 2011, 263n7) and to Clifford Geertz’s (1973b, 6) advocacy of ‘thick descriptions’ in anthropology (although, in Geertz’s use, the thick need not involve evaluation). Geertz in turn borrowed the phrase from Gilbert Ryle (2009b, 489; 2009c, 497). As Lipscomb (2021) indicates, G. E. M. Anscombe and R. M. Hare also played underappreciated roles in the renewal of interest in thick concepts.

⁴ This referential dimension is foregrounded notably in Hirsch (1993, 2013), Sider (2011), Cappelen (2013), Sawyer (2020a, c), and Campbell, O’Rourke, and Slater (2011)

to form a robust basis for their appraisal.

Irrespective of these metaphysical issues, however, there is a more basic reason to look beyond the referential dimension of concepts when appraising them: as Nietzsche notes, concepts do more than just turn the intellect into a pure mirror of the world.⁵ There may be a sense in which concepts are ‘representational devices’, as the conceptual engineering literature tends to describe them; but if so, they do more than represent.⁶ To take the full measure of a concept, we have to consider also what happens downstream of its application. What further reasons follow in the wake of the recognition that we have reasons to apply it? What does its applicability imply?⁷ A concept remains an empty label to its users unless it ‘locates its object in a space of implications’, in Wilfrid Sellars’s phrase.⁸ If the concept *F* were like an app that pinged when and only when presented with an *F*, the ping would be devoid of any significance for us unless we could infer something from it.

A concept’s merits therefore depend not just on whether anything corresponds to it in the world we inhabit, but also on what *follows* from its correct application, because that is what renders concepts, in the most literal sense, *consequential*: it is the primary way in which concepts make a difference to the rest of our thought and conduct. Two concepts that pick out the same set of objects, but associate them with radically different implications, are likely to differ also in their value to us. Concept appraisal should

⁵ ‘What if the intellect were a pure mirror? But concepts are more than that’ (70:8[41]). I follow Richardson (2020) in citing Nietzsche’s *Nachlass* by the last two digits of the year of the notebook in which the note occurs, followed by a colon, followed by the notebook number, followed by the note number in square brackets. Translations of Nietzsche’s texts are my own throughout, though I have consulted translations where available, and amended them only to bring them closer to the original.

⁶ A phrase popularized by Cappelen (2018, 3); in his usage, it allows for the fact that representational devices can act as expressive devices, however. But Mona Simion articulates the literature’s focus on the representational dimension when she writes: ‘Concepts, just like beliefs, are representational devices, their function is an epistemic one: to represent the world’ (2018, 923).

⁷ This two-faced model of the articulation of concepts that includes the consequences as well as the conditions of their application goes back to Michael Dummett’s (1973, 434) generalization of Gerhard Gentzen’s work on sentential connectives, and figures centrally in the inferentialist tradition (Brandom 1994, 2000, 2008; Dummett 1973; Kukla and Lance 2009). More recently, Jorem and Löhr (2022) have emphasized the importance of consequences of concept application for conceptual engineering.

⁸ See Sellars (1958, §107).

accordingly be sensitive to the reason-giving as well as to the reason-guided aspects of concept application. Inferentialists and referentialists about conceptual content may debate whether these inferential consequences should be regarded as directly constitutive of a concept's content or as indirectly associated with it. But we need not take sides on this metasemantic issue for our purposes. For, either way, it should be common ground that the concepts we actually possess systematically *co-vary* with the inferences we think we can appropriately draw.⁹ And certainly, it is only once we consider also what follows from concept application that we stand a chance of appreciating the effects concepts have even further downstream of their application, *via* their inferential consequences: the expressive functions they thereby discharge, for instance, or the needs they meet, or the concerns they help satisfy.¹⁰ A reasonably comprehensive picture of the respective merits of living by different concepts should encompass their wider impact on human affairs, and be sensitive to what concepts do for us *by* enabling us to refer and think about certain things.

This is especially true of thick normative concepts. For once people structure their affairs in terms of such world-guided concepts and become responsive to the action-guiding reasons they advert to, the concepts can end up closely dictating what people should do. Iris Murdoch evocatively calls this ‘the siege of the individual by concepts’ (2013, 31): if the applicability of those concepts is conclusively determined by empirical observation while their normative implications are non-negotiable because built right into the concepts, this can leave one feeling beleaguered by the concepts one uses, bereft of any room to reasonably dispute that they apply or that their applicability has certain normative implications.¹¹ It is therefore with these thick normative concepts that the mindware metaphor has most purchase—they really are codes of conduct, tightly linking certain worldly inputs to certain normative outputs.

⁹ See Marconi (1997) for a notably ecumenical account that supports this conclusion while accommodating both inferentialist and referentialist views within a dual aspect theory of lexical competence.

¹⁰ For a battery of arguments as to why we should look beyond reference when thinking about moral concepts, see Sinclair (2018).

¹¹ Whether one accepts this of course depends on how one analyses thick concepts; see Roberts (2013), Väyrynen (2013), and Eklund (2017, 88–93, 168–91) for discussions of the various analyses on offer. I follow Williams (1995i, k, m, o, 2005h, 2011, 2021) in my conception of thick concepts.

As a result, thick normative concepts also offer particularly effective tools by which to influence people. This helps explain why authoritarian governments tend to take an interest not just in the conclusions their citizens reach, but also in the very concepts with which they approach questions. It is perennially tempting for such governments, for example, to promulgate thick conceptualizations of *legitimacy* such that, for citizens who live under those governments, the only reasonable conclusion to be drawn is that their government is legitimate. And this is but a particularly significant illustration of a wider phenomenon that has led a string of observers to note the political dimension of questions of conceptualization or definition: ‘to choose a definition is to plead a cause’ (Stevenson 1944, 210); ‘[d]isputes over appropriate definitions are thus political conflicts’ (Sederberg 1984, 94); ‘definitions are a form of advocacy’ (Chesebro 1985, 14); ‘the choice of definitions is always political’ (Schiappa 2003, 68).

In practice, of course, the amount of influence achievable merely by disseminating certain concepts is easily overstated. Edward Bernays (1969) and other pioneers of propaganda and public relations may have claimed to possess the power to ‘engineer the consent of the governed’, but engineering consent by tampering with concepts has turned out to be a great deal harder than they initially made it sound.¹² Recent advances in digital technologies have rekindled concerns on this front, as they seem to make it unprecedently easy to shape how people think by filtering what they see.¹³ But we should be wary of claims made about the power of these technologies by those who conceive of it in the starkly simplified manner of conspiracy theories, or who have an interest in exaggerating it. Overstating the power of elites to determine how people think is, after all, a hallmark of what Richard Hofstadter (2008) called ‘the paranoid style’ in politics.

It may be felt that there is a more principled problem with this concept-centric way of manipulating people’s thoughts, moreover: that the beliefs people eventually arrive at remain underdetermined by the concepts people employ. If the aim is to instil the belief

¹² For a historical overview of the birth of public relations and Bernays’s role in it, see Tye (1998). On the use of propaganda to engineer consent, see Herman and Chomsky (1988), Handelman (2009), and MacLeod (2019).

¹³ For a recent exploration of the idea that digital technologies can shape how we think by filtering what we see, see Susskind (2018).

that x is F , then promulgating concept F instead of directly instilling the target belief seems to leave open the possibility of coming to the opposite conclusion, namely that x is *not* F . To manipulate people's concepts is to remain at one remove from their beliefs, leaving people just the degree of freedom they need to frustrate efforts at manipulation.

But although it is importantly true that what beliefs we form depends on more than just on what concepts we employ, it would also be implausible to deny concepts any influence on belief formation: surely, we should reckon with the subtle effects of *framing*, whereby the terms in which an issue is framed shape the judgements reached. As José Bermúdez (2021) has recently argued, such 'framing effects' are ubiquitous; and sensitivity to framing is not necessarily irrational: what concepts we frame our thoughts in can quite properly affect what reasons we take ourselves to have. Not all reframing is fraudulent relabelling.

Again, framing effects are particularly pronounced with thick normative concepts. Just because these concepts make normative issues turn on empirical observations that are hard to argue with, they are particularly effective at steering people more or less inexorably towards certain beliefs. As David Wiggins points out, the features of a situation can leave users of a thick concept 'nothing else to think' but that the concept applies, and hence nothing else to think but that its normative consequences apply with it.¹⁴

This means that the decisive work is often done already long before the moments of deliberation and choice, by what Murdoch calls the conceptually informed 'work of attention'. Becoming aware of those features of a situation that our concepts equip us to see continuously and imperceptibly 'builds up structures of value around us', with the effect that, when the time to consciously make a decision arrives, 'most of the business of choosing is already over', as Murdoch puts it; one is 'compelled almost automatically by what one *can* see' (2013, 36).

And yet thick normative concepts exert their subtle influence while giving concept-users the *impression* that they are freely making up their minds. That is why the power to channel attention towards certain features of a situation, or to frame an issue by casting it in certain terms, can be a particularly surreptitious form of power. 'When the concepts we are living by work badly', Mary Midgley observes, 'they don't usually drip audibly

¹⁴ See Wiggins (1990, 66).

through the ceiling or swamp the kitchen floor. They just quietly distort and obstruct our thinking' (1996, 1). That inconspicuous influence can be exploited. Promulgating a certain thick conceptualization of legitimacy instead of directly trying to instil the belief that the government is legitimate *seems* to leave open the possibility of judging that the government is not legitimate. In fact, however, the concept might be so closely world-guided in its application that one is left with 'nothing else to think' but that the government is legitimate: it clearly meets all the criteria for the application of a concept *the point of which* was to ensure that this government should meet them.¹⁵ Thick normative concepts can thus make particularly effective conceptual camouflage for attempts to manipulate beliefs.

At the same time, thick normative concepts also tend to be sociohistorically distinctive and local, so that there is a real question whether one needs to structure one's affairs in those terms. Just as we know that the smartphone could run an entirely different suite of apps, we know that we *could* think in different terms, since people have done and still do so.

This sense of alternatives, this hovering 'could', is more pronounced with thick normative concepts than with other concepts. *Thin* normative concepts, like *rational*, *good*, and *right*, for example, are far less world-guided in their application and may vary wildly in what they are concretely applied to; but at the level of the role they play in our reasoning, these thin concepts leave us less room for radical alternatives, since they seem to be concepts one is almost bound to gravitate towards by abstracting from the particulars of thicker judgements and generalizing over a variety of such judgements at once ('Are all of these different ways of going on *good*, or *right*, or *rational*?').¹⁶ And *purely*

¹⁵ For a systematic discussion of the political dimension of conceptual engineering and the liberal and democratic rationales for making it challenging to implement, see Queloz and Bieber (2022); on the risk of conceptual engineering being abused, see also Marques (2020), Ball (2020), Shields (2021b), and Podosky (2021).

¹⁶ On the lack of alternatives to thin concepts and the process of abstraction by which one arrives at them, see Williams (2011, 162) and Grönert (2016); Smyth (2020) understands the shift towards thinner concepts as driven by the idea that thin concepts have logical priority over thick ones and are better suited to the articulation of all-things-considered judgements. But see also Eklund (2017) for a thorough discussion of the intelligibility of variation even among such thin concepts and its implications for moral realism.

descriptive concepts, such as those of elemental chemistry or particle physics, are even less prone to give rise to the sense that we might think radically differently. They are, on the contrary, particularly apt to invite the idea that the right set of concepts is the one that faithfully mirrors the structure of the natural world we inhabit—and that if there is only one such world, there is only one right set of corresponding concepts.

With thick normative concepts such as *chaste*, *sinful*, *chivalrous*, *unkind*, *phoney*, *courageous*, *dishonourable*, *dignified*, *treasonous*, *respectful*, *rude*, *elegant*, *vulgar*, *kitsch*, *sublime*, or *creepy*, by contrast, the ‘one world, one right set of concepts’ model soon gives out. Though there is but one natural world, the social worlds we have lived in are many, and there are many more we could come to inhabit. To ask which thick ethical, political, legal, cultural, and aesthetic concepts we should use is to ask which social world we want to live in. We may be more closely guided by how the world is in applying these concepts, but we have correspondingly more freedom *not* to cast our thoughts in these particular terms at all. Thick normative concepts thus do more to predetermine the run of things than thin or descriptive concepts, while being at the same time under more pressure to assert their place against alternatives. This combination of features ensures that the thick normative concepts that lend different social worlds their distinctive character render the question of their *authority* particularly apposite.

Murdoch also registers this when she adds, after highlighting the compelling force of ‘what one *can* see’: ‘This does not imply that we are not free’ (2013, 36). It implies, rather, that our freedom is exercised not only in the choices we make in view of what we *can* see, but also in our ability to determine *what* we can see. Hence the importance of making the effort to *look again*, which Murdoch illustrates with the mother-in-law who, upon realizing that her view of her daughter-in-law as *vulgar*, *undignified*, *noisy*, and *tiresomely juvenile* may merely be an expression of her own jealousy, puts in ‘the work of attention’, and discovers her daughter-in-law to be ‘not vulgar but refreshingly simple, not undignified but spontaneous, not noisy but gay, not tiresomely juvenile but delightfully youthful’ (2013, 17). We remain free to break old habits of thought and refocus our attention, not merely by redeploying it, but by redeploying it in terms of different concepts.¹⁷ The freedom we have in determining how to live, and which among the

¹⁷ On the plasticity of our ‘habits of thought’ and how they can be either insufficiently or overly plastic, see Delacroix (2022, 5, 10, 59–88).

concepts we live by to apply, extends also to the question of which concepts to live by.

We need not look as far as the gulfs between social worlds for differences in the concepts we live by to start to matter, moreover. Even small differences in conceptualization can entrain large differences in practice. This is what motivates Philip Pettit (1997) and Quentin Skinner (1998) to advocate a shift from conceptualizing liberty as *non-interference* to conceptualizing liberty as *non-domination*, for example. Liberty as non-interference is secured just as long as other people do not interfere in one's affairs, while liberty as non-domination additionally requires that other people should not even possess the *capacity* to interfere in one's affairs on an arbitrary basis:¹⁸ the Roman slave whose benevolent master never interferes in his affairs is still being dominated by his master—his *dominus*—who could change tack at any moment.

This seemingly small difference in the conceptualization of liberty can spawn diametrically opposed views on any number of contemporary issues, such as how we should think about the large-scale harvesting of our personal data. If we conceptualize liberty in terms of non-interference, the large-scale harvesting of personal data does not count as a reason to think that our liberty is being undermined as long it does not interfere with the exercise of our will; but if we conceptualize liberty in terms of non-domination, the same data harvesting does count as a reason to think that our liberty is being undermined, since we whose data have been harvested depend on those who control the data not to use them against us. While the former conceptualization gives us no cause for alarm, the latter gives us every reason to protest.

The divergent ramifications of endorsing subtly different conceptualizations become particularly salient when we consider the requirements that our value concepts place on the design of new technologies. As a recent article in *Ethics and Information Technology* observes: ‘Politicians and engineers are increasingly becoming aware that *values* are important in the development of *technological artefacts*. What is often overlooked’,

¹⁸ This last qualification importantly allows that there can be interference without domination as long as the interference is non-arbitrary, i.e. constrained and justified. But it in turn invites the question whether this concept of liberty blinds us to the real costs in freedom involved even in non-arbitrary interference—a point I address in Chapter 9; see also Lane (2018) and Cueni (manuscript-c). The potential impact of data-harvesting on liberal democracy is explored in Zuboff (2015, 2019), Nemitz (2018), Macnish and Galliott (2020), and Vélez (2020).

however, ‘is that different *conceptualizations* of these abstract values lead to different design-requirements’ (Veluwenkamp and van den Hoven 2023, 1). Demanding that new technologies be shaped to our values may be a start, but it still leaves all the work to be done: notably, the work of deciding which conceptualizations to shape the technology *to*.

All of which brings us back to the question we started out from: what kinds of reasons are there for us to accept certain conceptualizations and their correlative patterns of reasoning rather than others? These would have to be reasons to *count* certain considerations *as* reasons for or against other considerations. A philosophical framework is required to help us to think about these reasons for reasons—they are the reasons we need to identify in order to decide which conceptualizations to adopt, adhere to, or abandon.

As Wittgenstein pointed out, however, criticizing or justifying a thought on the basis of other thoughts is one thing, criticizing or justifying a *way of thinking* is quite another.¹⁹ In justifying one thought by another within a certain way of thinking, we take the concepts structuring that way of thinking for granted. Once we step back and attempt to justify that way of thinking itself, however, entirely different types of considerations appear to be called for. While the concepts we use evidently play a role in determining which reasons to think something true or justified we are sensitive to, concepts cannot themselves be true or justified the way judgements, propositions, or beliefs are. We therefore cannot model concept appraisal on the more familiar business of assessing the veracity or warrant of individual judgements. We can give reasons within the practice of reasoning defined by our concepts, but it is a different challenge altogether to give reasons for a way of reasoning. Wittgenstein even wondered whether it was possible to ‘give a reason for thinking as we do’, or whether this would—incoherently—‘require an answer outside the game of reasoning’ (1979, §4). Can we give reasons for the way we reason? It can seem as if any such justification must presuppose what it is meant to justify if it is to

¹⁹ See Wittgenstein (2005, 180); he puts it in terms of the activity of thinking as a whole, but the fundamental point—that very different types of considerations are called for once one goes beyond justifying one thought on the basis of another within a certain way of thinking—holds in either case. I explore this theme in his work in Queloz (2016, 2017). A related distinction is drawn by Toulmin (1953, ch. 11) and Rawls (1955).

be accessible to the addressee.

And yet we evidently do sometimes give reasons for or against certain ways of reasoning, and manage to think critically *about* concepts rather than with them. There must be a way to make philosophical sense of this fact. That is not just an explanatory demand, but an *ethical* demand. We need some way of determining whether our concepts are helping us to live—whether we are using concepts that express and serve our concerns, or whether we are, as Nietzsche feared, ‘stuck in a cage, imprisoned among all sorts of terrible concepts’ (2005c, *Improvers*, §2). But how do we tell the difference? How can we critically ascertain that the building-blocks of our thoughts are not stumbling-blocks to our concerns?

My aim in this book is to develop a framework for concept appraisal. At the heart of this framework are *reasons for concept use*, a special class of reasons that are orthogonal to, and yet can underpin or undermine, the reasons for action and belief that figure in our deliberations. To adopt a concept is to become sensitive to the reasons that immediately guide and flow from its application. But we can go one step further and ask for reasons to reason in these terms. This is to demand reasons for reasons—second-order reasons to use the concept and be responsive to its concomitant first-order reasons.

As Wittgenstein’s puzzlement reminds us, it is a remarkable fact that we can give and ask for reasons for the way we reason *from within* our practices of reason-giving. This is the accomplishment of the concept of a second-order reason: it makes it possible to subject to critical scrutiny the very concepts and first-order reasons that make up the conceptual architecture we inhabit. By acquiring the ability to think in terms of reasons for reasons, we become able to sound out and renovate the house of reason from within.

Pursuing the question of how best to do this will lead us deep into what is increasingly emerging as a field in its own right—a field which, following Alexis Burgess and David Plunkett, I will refer to as ‘conceptual ethics’.²⁰ Though it may now be in the ascendant,

²⁰ For overviews of this emerging field, see Burgess and Plunkett (2013a, b) and Cappelen and Plunkett (2020). Recent contributions congenial to the approach pursued here include Miller (2010), Plunkett (2015, 2016), Fredericks (2018, 2020), Goetze (2018, 2019), Latham, Miller, and Norton (2019), Haslanger (2014, 2020a), Thomasson (2020b, 2022), Isaac (2021), McPherson (2020b, 49–52), McPherson and Plunkett (2020, 2021), Nado (2021b), Smithson (2020), Shields (2021a, b, c), Lau (2022), Jorem (2022), and Santarelli (2022). An earlier manifesto for this type of philosophical inquiry is

conceptual ethics is not new. Philosophers have long been in thrall to the ideal that when we are being rigorous in our reasoning, we should not listen to any old reason flowing from any old concept, but should probe how far the reasons suggested by our concepts can be shored up by reasons validating the rational authority of these concepts.²¹ That is part of what it means to reason for oneself rather than to rely on guidance from elsewhere, and hence part of what marks the difference between rational autonomy and rational tutelage that forms the defining contrast of the Enlightenment according to Immanuel Kant.²² We should not use a concept merely out of awe, or merely out of habit—intuition should be grounded in argument, suggestiveness in discursiveness. As Kant insisted, we should query *by what right* our concepts tell us how to think.²³

This demand that we should always be able to offer reasons for using the concepts we use may have received its purest expression in Enlightenment rationalism. Yet, to this day, we are familiar with the philosophical interlocutor who greets justifications such as ‘Because it’s a *human being*’ with: ‘So what?’; or ‘Because it’s *natural*’ with: ‘Why does that

formulated by Carruthers (1987), who advocates an approach that is similar in spirit to the one advocated here, as does Schiappa (2003), who offers rhetorical-cum-philosophical analyses of several case studies to illustrate the politics of definitions. Pettit (2020, 353n23) explicitly identifies his earlier work as an example of conceptual ethics.

²¹ Intimations of that ideal can be gleaned already from Plato’s insistence, in the *Theaetetus* (201c–210d), that a claim needs a *logos* (a reason) to count as knowledge, and from his call, in the *Protagoras* (356d–e), to turn practical reasoning into a *techne*—a ‘science’ or ‘measurement system’—that promises to shelter people from *tyche*, the unpredictable play of fortune, by giving them more control over whether their lives go well. See Nussbaum (2001, ch. 4).

²² In his essay ‘*Was ist Aufklärung?*’, Kant defines Enlightenment as humanity’s emergence from its self-imposed tutelage, and characterizes ‘tutelage’ or *Unmündigkeit* as the inability to use one’s own reason without another’s guidance (1900–, WA, AA 08: 35.1–3). On a related contrast to be found in Islamic philosophy between *taqlid*, the uncritical acceptance of authority, and *ijtihād*, judgment based on independent effort, see Adamson (2022).

²³ Kant’s transcendental deduction of the pure categories of the understanding (A84–130/B116–169), which he offers in answer to this ‘*Quid juris?*-question, is perhaps the most venerable example of an attempt to formulate reasons by which to ground the legitimacy of concepts—Christine Korsgaard’s self-consciously echoes Kant’s question in *The Sources of Normativity* when she asks after ‘the right of these concepts to give laws to us’ (1996, 9). Yet, by Kant’s own lights, the strategy of transcendental deduction is only available for a handful of special concepts—the categories—and cannot be generalized to the sociohistorically local and thick normative concepts I focus on here.

give me a reason?' This is not to question the applicability of the concept *human being* or the concept *natural*; it is to question the *authority* of these concepts—to ask for reasons to treat those concepts as giving one reasons. By demanding reasons to reason in certain terms and heed the reasons they advert to, we challenge the authority of concepts. To vindicate the authority of concepts, we need reasons for reasons.

Traditionally, such reasons for reasons have been sought in timeless rational foundations capable of authenticating one set of concepts as absolutely and definitively best: the concepts corresponding to Platonic Forms or the Mind of God, perhaps, or the concepts dictated by natural law or universal reason, or the conceptual scheme whose structure is the mirror-image of the structure of reality. The normative expectation that one should be able to offer reasons for reasons thus forms a potent driver towards foundationalist theory-building.²⁴ As Bernard Williams observes,

the most powerful thing that drives toward theory is this idea that you must give a reason for a reason. If I say: But it's an animal, they say: Why is that a reason? I've got to give a reason for that reason, in the end. That's why I end up with foundations. (1999, 251)

The effect of this expectation that one should be able to give reasons for reasons that can ultimately be anchored in rational foundations, however, is that concepts deprived of such anchoring are demoted to mere conceits, to be cast aside by the rational thinker. And if the conclusion reached is that there are in fact no timeless rational foundations to be had—at least not widely enough to support a *generalized* foundationalism—the same expectation that we should derive the set of truly authoritative concepts from such foundations creates a sense that something crucial is *lacking*, and results in alienation from one's concepts. Absent timeless rational foundations, the realization that there are other concepts we could use, and that there is no neutrally identifiable reason to prefer the concepts we happen to have over alternatives, encourages indiscriminate disengagement from all our concepts.

Attempts to reconcile a reflective sense of the contingency of our concepts with full-blooded confidence in those concepts have tended to err in the opposite direction and license the undiscriminating *acceptance* of whichever concepts we inherited. Fuelling this

²⁴ As comes out in Radzik's (2000) discussion of different foundationalist theories of normative authority, for example.

reconciliation, usually, is some kind of holism about our conceptual repertoire, which comes in two versions. On one version, it does not matter which concepts we use, as long as the judgements we form with them are true and the concepts combine well enough to form a functional whole. Agonizing over which concepts to use then seems like agonizing over the weave pattern of a fishing net: such a net could be composed of triangles, squares, hexagons, or intricately combine all of these, but it is immaterial which pattern we use as long as the net's overall integrity is preserved. The choice of which web of concepts to use may be constrained at the edges by merely pragmatic considerations—there are some webs of concepts that human beings, with their limited cognitive and perceptual capacities, are incapable of deploying effectively—but otherwise, the choice remains rationally undetermined: it is voluntaristic or arbitrary. As Huw Price describes the view: 'Not only is language less of a prison than philosophy usually imagines ... we can put the walls wherever we like!' (2018, 469).

In the second version of this holistic view—associated, notably, with functionalist holism in anthropology and with communitarianism in social and political philosophy—the holism is given a more explicitly functionalist basis: a society's conceptual repertoire is pictured as a harmoniously interlocking whole that has organically grown out of a particular way of life, and has over time become adapted to that way of life. This makes it the best conceptual apparatus *for* that way of life. Displace one concept, and you diminish the functionality of the whole.

Both forms of holism manage to rid themselves of the hankering after timeless rational foundations: they are genuinely *non-foundationalist*. But the result, in either version, is that indiscriminate rejection gives way to undiscriminating acceptance. The result is an 'enthusiasm for the folk-ways' that has been called 'the continuation of Hegelian conservatism by other means' (Williams 2021, 278).

My aim in this book is to develop a theory of reasons for concept use that does without foundations, and yet still gives us a basis on which to discriminate between concepts and ground our confidence in reasons, so that we may escape the trilemma between foundationalism, indiscriminate ironism, and undiscriminating holism. I try to do without the idea that there are timeless rational foundations from which one set of concepts could be authenticated as absolutely best—indeed, I try to do away with the very idea that such timeless rational foundations, whether available or not, would be

desirable. At the same time, I hold on to the idea that it really does matter which concepts we use, because which true judgements we are capable of forming, and which reasons we are responsive to, significantly shapes how we conduct our affairs. What is needed to occupy this middle ground is a framework for concept appraisal that does not require timeless foundations while still enabling us to discriminate between concepts on a case-by-case basis instead of only considering the web of our concepts as a whole.

One influential way of appraising and improving our conceptual apparatus piecemeal without relying on foundations has been to fasten on how messy and defective our inherited concepts appear when measured against the formal ideal of a tidy theory: many of our concepts are imprecise, vague, indeterminate, inconsistent, and incoherently related to other concepts. By fixing these defects, philosophers can move towards concepts that are more precise, determinate, consistent, and coherent. The resulting concepts may not timelessly and definitively be the best ones, but their precision, determinacy, consistency, and other theoretical virtues promise to guard against the dangers inherent in slovenly thinking.

Yet I want to resist this view as well—or rather, I want to put it in its place and expose what it leaves out. Faced with theoretically virtuous concepts recommending something that radically conflicts with what our untidy but entrenched concepts lead us to think, it becomes a real question why we should care so much about theoretical virtues as to overturn concepts that have at least as much force with us. To answer that question, a more comprehensive approach is needed, one that can situate the importance of theoretical virtues within a wider picture of what we want from our concepts, and make sense of how even theoretical vices can be virtues in other respects.

To this end, the book develops a *needs-based* approach to concept appraisal: it proposes to appraise predicates by predicaments, determining which concepts we should use by identifying *the concepts we need*. Instead of simply requiring us to conform our concepts to an order of things about which we do not have a say, or to a tidy theoretical structure, it humanizes the standard to which concepts are answerable, turning that standard itself into a function of human concerns. On the resulting view, we are the ones who *authorize* our concepts, in both senses of that useful term: we are their authors, and we lend them authority, for it is by tying in with human concerns that concepts earn their keep.

At the same time, the standard of concept appraisal should not *just* be a function of human concerns, but also of how the world is and what capacities we bring to it, leaving us firmly constrained by reality in what kinds of concepts we can have reason to use. What concepts we need is a function of the demands our situation makes on us and our capacities as a result of our pursuing certain concerns. Conceptual ethics is a three-way negotiation: what our concepts should demand of us depends on what the world demands of us given what we demand of the world.

Accordingly, the guiding intuition of the needs-based approach is that the *value* of a concept lies in the way it proves *of value to us* by meeting our conceptual needs—our needs for certain concepts that result from the conjunction of our concerns, capacities, and circumstances. These needs are different from the goals, aims, and purposes that figure so prominently in broadly functionalist approach to conceptual ethics and engineering.²⁵ Our conceptual needs are something we may or may not be cognizant of. They are not something we consciously pursue, like goals, aims, or purposes, but rather grow out of the way the concerns we pursue interact with the capacities with which and the circumstances in which we pursue them. As I will argue, it is when concepts align with the conceptual needs we have in virtue of concerns we endorse that they are authoritative.

Dauntingly complex as these interactions are, I will suggest that we can render them philosophically tractable by constructing what I call a *need matrix*: a three-pronged interpretative model of how the concerns, capacities, and circumstances of concept-users combine to generate a conceptual need. Such a need matrix can be used to determine what kind of concept best meets that need. Like the matrix from which printing type is cast, the need matrix acts as a mould from which to cast fitting building-blocks of thought.

The first of the two main theses of the book, then, is that if our minds are moulded by our concepts, our concepts should be moulded by our needs. For it is in our conceptual needs, which grow out of the way the concerns we pursue interact with the capacities and circumstances in which we pursue them, that we find reasons for concept use. In light of those reasons, concepts will emerge as authoritative when they align with the needs we

²⁵ See, e.g., Burgess and Plunkett (2013a, b), Brigandt and Rosario (2020), Thomasson (2020a, b), Nado (2021b), Haslanger (2020b), Simion and Kelp (2020), Riggs (2021), and Jorem (2022).

now have, in virtue of concerns we identify with and would endorse upon reflection.

This needs-based view of conceptual authority thus breaks with philosophy's traditional quest for the concepts that are absolutely and definitively best. Yet despite this, I will argue that this approach allows us to reconcile ourselves to the contingency of our concepts and avoid indiscriminate alienation from our concepts.

At the same time, the needs-based approach also avoids the opposite danger—the undiscriminating acceptance of whatever concepts we inherited. Appraising concepts according to our conceptual needs enables critical reevaluations of our conceptual inheritance. It promises to give us a nuanced and case-specific sense of which conceptualizations we have most reason to operate with. As Chapters 8 and 9 will illustrate, it even empowers us to assess competing proposals for how to understand such contested concepts as *voluntariness* or *liberty*, thereby providing some basis for distinguishing between authoritative definition and conceptual gerrymandering.

The second main thesis of the book is that sometimes, concepts that conflict, or that exhibit other theoretical vices such as vagueness or superficiality, are just what we need. This is what sets off the needs-based approach to conceptual authority from its closest non-foundationalist rival, the tidy-minded pursuit of conceptual authority through theoretical virtue, which models its understanding of what makes a good concept on the virtues of a scientific theory.

By developing a way of vindicating the authority of concepts without invoking the authority of theoretical virtues, the needs-based approach indicates a different way of practising philosophical reflection on our concepts, one that relies less on theory-construction and the realization of theoretical virtues, and more on the particular psychological, social, and institutional facts on the ground. If we consider not what concepts are timelessly and definitively best, but what concepts we now need, we can identify the proper remit of efforts to tidy up our conceptual repertoire, and thereby come to discriminate between helpful clarification and hamstringing tidy-mindedness.

If the primary theme of the book is reasons for reasons, its secondary theme is therefore how the demand for such reasons can, depending on how one conceives of those reasons, press towards foundationalist theory-building and the tidy-minded pursuit of theoretical virtues, and why that pressure should sometimes be resisted. The question of what form reasons for reasons should take thus bears directly on the question

of what form philosophy should take.

My thoughts on these issues have been shaped by two figures whose influence pervades the book. One is Friedrich Nietzsche, who is the figure in the Western philosophical canon that most dramatically embodies a sceptical readiness to question the authority not just of concrete institutions and individuals, but of the ideas they abide by. While Nietzsche's scepticism towards the authority of concepts is Kantian at root, his distinctive elaboration of that scepticism into what he billed as a 'real critique of concepts' (85:40[27]) is all the more powerful for being piecemeal and focused on culturally distinctive constellations of concepts rather than on the authority of our reasoning faculty as a whole. It is his use of philosophy as piecemeal cultural critique that makes Nietzsche's thought so modern and marks it off from Kant's Enlightenment universalism.²⁶ Nietzsche explores the ramifications of the realization that a culture's trajectory, and *a fortiori* the trajectory of an individual life, are shaped by the concepts people recognize as authoritative. If conceptual ethics has an avatar, it must be Nietzsche.

The other figure is Bernard Williams, whose *Ethics and the Limits of Philosophy*, in particular, shaped my views on the place and merit of parochiality and tidiness in our thinking. Williams is routinely cast and taught as a purely negative and destructive philosopher, who cleverly chipped away at others' honest attempts at philosophical theorizing, but offered little by way of an alternative. I take the view I articulate in this book to be indicative of the more positive conception of philosophical reflection that underpins his work. By building a framework for conceptual ethics that develops, fuses, and harnesses various insights scattered across Williams's contributions to metaphilosophy, epistemology, ethics, metaethics, political philosophy, and the theory of action, we become equipped to appreciate how much of Williams's own work was itself a methodologically cohesive and constructive exercise in conceptual ethics—in line with his declared conviction that 'the task which provides the principal aim of all moral philosophy' is 'the ethical understanding of the ethical', the task 'of truthfully understanding what our ethical values are and how they are related to our psychology, and making, in the light of that understanding, a valuation of those values' (1995a,

²⁶ As noted also by Huddleston (2019, 171).

578).²⁷ If Williams's work can inform this book, it will emerge, it is because he was himself, alongside Nietzsche, a paradigmatic practitioner of conceptual ethics.

The book can be divided into four parts. The first part (Chapters 1–2) sets up the discussion by introducing the guiding question, clarifying its terms, and laying out the various conceptual tools we will need to answer it. Its main themes are the very idea of conceptual authority and how it differs from the normativity of concepts; the powers of concepts that motivate raising the question of their authority; our capacity to adopt the autoethnographic stance towards our concepts; the distinction between engaged and disengaged concept use; the confusions that lead to objectionable forms of relativism; the different ways in which our concepts are contingent; the way our confidence in our concepts can be undermined by reflection and cause us to lose knowledge; but also the possibility of acquiring metaconceptual knowledge that certain concepts are the right ones for us.

The second part (Chapters 3–4) surveys the different answers that the question of the authority of concepts has received in the past. Discussing their shortcomings motivates the development of an alternative account and conveys a sense of the features it should have: it should generalize to thick normative concepts; it should not result in indiscriminate disengagement from our concepts; and it should not license the undiscriminating acceptance of our concepts. To this end, I suggest, the picture of our conceptual apparatus as something harmonious, largely tensionless, and inherently static must be replaced with a kaleidoscopic picture on which our conceptual apparatus is tension-ridden and dynamic; and the critical leverage of local needs must be harnessed by recognizing that the contingency of our concepts extends to the standards these concepts must meet. This still leaves one direct rival to the approach I aim to develop, however: the tidy-minded approach that seeks conceptual authority by eliminating theoretical vices such as vagueness, superficiality, and tensions in our concepts. I explore

²⁷ Not only Williams's critique of the particular forms that certain important moral concepts take within 'the morality system' and his critique of deepened conceptions of the voluntary can be thought of as exercises in conceptual ethics, but also his more constructive advocacy of certain conceptions of agency and responsibility, shame, regret, reasons, thick ethical concepts, and virtue-ethical concepts, including notably the 'virtues of truth'. As he also writes: 'our conceptions of freedom, responsibility, and blame are often not what they seem, and are variously exaggerated, self-deceiving, sentimental, or vindictive (epithets which themselves, it should be noticed, largely belong to an ethical vocabulary)' (1995a, 578).

these different theoretical vices and offer reasons to be critical of the tidy-minded approach as a general account of conceptual authority.

The third part (Chapters 5–7) lays out the needs-based approach to concept appraisal I advocate in this book. Using a little-known debate between Ronald Dworkin and Bernard Williams as my springboard, I first introduce the basic idea that our concepts should make contact with the human concerns motivating their use to begin with. This brings out more sharply what the tidy-minded focus on theoretical virtues misses. But I argue that our concerns alone are insufficient to determine which concepts we should use. A more complex framework is required, which I go on to develop by introducing the notions of a conceptual need, of the expressive character of concepts, of needfulness conditions, of the pointfulness of concepts, and of need matrices. These notions combine into a powerful framework for needs-based concept appraisal. I then show how this allows us to answer the authority question without crowding out other kinds of reasons that transcend preoccupation with human concerns or with the instrumentality of concepts. On this basis, I articulate a needs-based conception of conceptual authority, on which concepts are authoritative if and to the extent that they meet the conceptual needs we have in virtue of concerns we identify with and would endorse upon critical reflection. I defuse the worry that this conception yields the wrong kinds of reasons, and explore in what sense it still leaves room for concepts to be valuable in their own right, independently of how they serve our concerns.

Finally, the fourth part (Chapters 8–9) illustrates and refines the account by considering its application to the specially complex cases of the concepts of voluntariness and liberty, where it emerges that superficial and conflicting concepts sometimes serve us best. These case studies not only further illustrate the approach, but also underscore several further insights it yields: that sometimes, powerful concerns can distort concepts out of the shape in which they best serve the balance of our concerns; that the very conflict and heterogeneity of human concerns can itself generate reasons to use certain concepts rather than others; and that there is a place in liberal democratic politics for the tidy-minded pursuit of theoretical virtues.

The hurried reader seeking to understand the mechanics of my view without much by way of motivating background, contrast foils, and detailed applications should focus on Chapters 1–2 and 5–7. By the end of the book, I hope to have shown that the way of

thinking about concept appraisal we need is itself one that appraises our ways of thinking by our conceptual needs.

CHAPTER ONE

The Authority Question

1.1 Dworkin's Challenge

'Nothing is easier than composing definitions of liberty, equality, democracy, community and justice', Ronald Dworkin once wrote, 'but not much, in philosophy, is harder than showing why these are the definitions that we should accept' (2001a, 90). If we cast our reasoning in certain terms, it may well follow that we have reason to draw certain conclusions; but it remains open to us to ask why we should recognize the authority of those terms to begin with. Why should we place ourselves under the sway of these conceptualizations if alternatives are available that would lead us to different conclusions? This is what Dworkin identifies as the real challenge for composers of definitions: the challenge of answering what I shall call *the authority question*.

We are familiar with the idea that we can question the authority of individuals or institutions: why should this individual, or that institution, be granted authority over our lives?¹ Yet the same question can fruitfully be extended to concepts. Max Horkheimer

¹ Theorists of who have elucidated the notion of authority notably include Weber (2019, ch. 3), Horkheimer (1987), Arendt (1956, 1958), Jüvenel (1963a, b), Kojève (2014), Friedrich (1958, 1972), Eschenburg (1976), Green (1988), Barber (2010), Christiano (2020), and Kletzer and Renzo (2020). The account of authority that has been most influential within analytic philosophy in recent years is Raz's (1979, 1986, 1995, 2009), and I discuss its relation to the conception I advocate here in Chapter 7. I take the notion of *conceptual* authority at issue in the authority question to have precedents in Dorsey (2016, ch. 1), McPherson (2018, 2020a), Plunkett (2020), and Wei (2022); related notions are also invoked in

described authority as *bejahte Abhängigkeit*—deliberate dependence.² When we place ourselves under the sway of a concept, we willingly make our judgements and our actions dependent on the reasons the concept adverts to and the way of thinking it prescribes. We can therefore meaningfully ask *why* we should grant a concept the authority to shape our thought and conduct. This is not merely to question the authority of a concept on a particular occasion ('Does the fact that x is F really matter here?'). It is to ask why we should ever think in those terms (' F -ness') at all.

What the authority question invites us to identify are, in the first instance, not reasons for action or reasons for belief, but *reasons for concept use*; not reasons operative *within* a practice of reason-giving, but reasons *for* a practice of reason-giving, i.e. reasons to adopt or adhere to a concept and be disposed to recognize certain considerations *as* reasons.

This is notably different from asking what justifies individual thoughts and actions *given* the concepts one uses: my reason for thinking that x is F might be what I take to be the fact that x is G , and I might pursue the chain of reasons further in that direction by asking what reasons I have for thinking that x is G . But the authority question leads us, not further along the chain of first-order reasons, but along a different chain that is orthogonal to the first, towards higher-order reasons: it raises the *metaconceptual* issue of what reasons I have to *count* the fact that x is G *as* a reason for thinking that x is F . In other words, the authority question asks for second-order reasons to think in terms of certain first-order reasons. As this characterisation of conceptual authority in terms of reasons for reasons suggests, I understand conceptual authority as a gradable and comparative notion: I might have more reasons, or better reasons, to use one concept rather than another.

The relevant notion of conceptual authority must be distinguished from a related notion in the vicinity that has received rather more attention, namely what Robert Brandom calls that 'special sort of *authority* one becomes subject to in applying concepts' (1994, 9), that 'rulishness' or 'normative bindingness' (1994, 10–11) whereby one

Stampe (1987), Johnston (2001), Ridley (2005), Chang (2009), Hayward (2019), Wodak (2019), and Smyth (2018, 2022). For a disambiguation of various senses in which *reasons* can have authority, see Hampton (1998b, 85–93).

² See Horkheimer (1987, 24, 46). Jüvenel (1963a, b) also thinks of authority in terms of internal assent rather than in terms of obedience.

becomes liable to assessment as to whether one applies a concept correctly or not. This is the authority that a concept has just insofar as it functions as the standard for assessments of the correctness of reasoning involving that concept. Whether this form of conceptual authority as normative bindingness even exists, and what its nature and source might be, has been the subject of much debate in the literature on the normativity of concepts.³ But the aim has been to elucidate that authority in terms meant to apply to *all* concepts rather than to help us discriminate *between* them. As Rebecca Kukla puts it, the aim has been to explain how we could possibly ‘get inducted into normative space in the first place’ (2000, 162). The authority question, by contrast, asks for reasons to recognize a particular concept’s normative bindingness.⁴ Why, given all the concepts we could be using instead, should we place ourselves under its sway?

We can mark this contrast by distinguishing between authority *of* use and authority *in* use. Questioning a concept’s authority of use problematizes the *legitimacy* of a concept’s power over us: *by what right* does a concept tell us how to think?⁵ Questioning a concept’s authority in use problematizes the *normativity* involved in a concept’s exercise of that power: *how is it possible* for the claims a concept makes on us to be binding? When Brandom speaks of the ‘authority’ of concepts, or their ‘grip on us’, he means their authority *in* concept use.⁶ When I speak of their authority or their hold on us, I mean

³ See Ginsborg (2018) for an overview of the debate over how to understand the normativity of concepts and what norms, if any, guide concept use and underlie assessments of it.

⁴ I thus agree with Burgess and Plunkett (2013a, 1095–6) in distinguishing the issue of whether to use a concept at all—which is the issue I am concerned with—from the issue at the heart of the literature on the normativity of concepts, of how to use a concept and what norms, if any, guide concept use.

⁵ In Kant’s terms, the authority question is thus a ‘*Quid juris?*-question, which his transcendental deduction of the pure categories of the understanding (A84–130/B116–169) is explicitly offered in answer to. His concern with legitimacy is reflected also in his self-conscious redeployment of the legal term ‘*Deduktion*’, which, in its original context, designates the legitimization of property claims by historical derivation.

⁶ ‘The most urgent question for Kant is how to understand the rulishness of concepts, how to understand their authority, bindingness, or validity. It is this normative character that he calls *Notwendigkeit* (necessity)’ (Brandom 1994, 10). ‘For Kant the question is ... how to understand [concepts’] grip on us: the conditions of the intelligibility of our being bound by conceptual norms’ (2009, 33). He uses these phrases in the same sense in Brandom (2000, 80, 164; 2002a, 22; 2019b, 9). Rouse (2015) similarly focuses on authority in use, i.e. on accounting for the binding force of concepts.

their authority of concept use.

The authority question as I understand it arises independently of how we understand authority in use. In particular, the intelligibility of Dworkin's challenge does not require concepts to be normative in the strong sense that involves being consciously guided by explicitly represented norms as one applies a concept. It only requires the idea that our concepts in fact exert power over us, and that different concepts would exert power over us differently. Even if this power manifests itself in matter-of-factual dispositions to conceptualize the world in certain terms and reason along certain lines, without conscious guidance by norms, this still leaves room for concepts to be normative in the sense that one is *subject to normative assessment* by third parties as one applies a concept, thereby finding an independent basis outside the thinking individual for the distinction between correct and incorrect concept application.⁷

But however exactly normativity comes to be involved in a concept's exercise of its power, my focus here lies on the normatively prior issue of the legitimacy of that power. The question raised by Dworkin's challenge, in other words, is not how a given concept binds us, but why we should let it.

1.2 The Power of Conceptual Architectures

If it makes sense to extend the authority question to concepts, it is because, like powerful individuals and institutions, concepts exert various forms of power over us: they not only steer our attention, channel our emotions, direct our imagination, and marshal our memories; they also govern our reasoning by prescribing particular patterns of thought, telling us what to care about, what to count as a reason for what, and which inferences to draw.⁸

It is often said that concepts are the building-blocks of thought; but for creatures whose lives are as pervasively shaped by thoughts and ideals and rules as ours, that is tantamount to saying that they are the building-blocks of our lives.⁹ We inhabit a

⁷ A point emphasized by Hlobil (2015).

⁸ These different aspects of the power of concepts are well described in Diamond (1988), Eberhart (2004), Fricker (2007), Haslanger (2012, 2018, 2020a), Nguyen (2020), and Kaeslin (Manuscript).

⁹ Daston (2022, 1) vividly makes a similar point about rules, and what holds for rules holds a fortiori for

conceptual architecture as much as a material one. Many of the objects that surround us only get their meaning and point from being embedded in behavioural schemas that are organized by concepts.¹⁰ Not only do concepts help us organize an antecedent reality. They also contribute to building the social world we inhabit—something Iris Murdoch expressed when she wrote that a concept is ‘less like a moveable and extensible ring laid down to cover a certain area of fact’, and more like something that determines the ‘*Gestalt*’ or ‘configuration of the world’ (1956, 40–41; 54–55).¹¹

This becomes most evident when we consider societies structured by concepts that are very different from our own. It is then especially easy, as Williams observes in a similarly architectural metaphor, to ‘see their judgments as part of their way of living, a cultural artifact they have come to inhabit (though they have not consciously built it)’ (2011, 163).¹² Just as the material edifices we move in enable, guide, and constrain the movements of our bodies, the conceptual edifices we inhabit enable, guide, and constrain the movements of our minds. This can be as restricting as it can be empowering. In keeping with the image of a conceptual architecture, John Steinbeck renders the ambivalence thus: just as we might stumble down narrow dark alleys of thought, we might walk up the stairs of our concepts to emerge ahead of our accomplishments.¹³

Our conceptual architecture *enables* us in that it allows us to think and do things that we could not think or do without it. New concepts can make the invisible visible, disclose

the concepts in terms of which they articulated: ‘Rules fix the beginning and end of the working day and the school year, direct the ebb and flow of traffic on the roads, dictate who can be married to whom and how, situate the fork to the right or the left of the plate, score the runs and walks of baseball games, tame debate in meetings and parliaments, … specify who can vote and when, … and order the rites of birth and death. And these are just examples of explicit rules Add implicit rules, and the web becomes so densely woven that barely any human activity slips through the mesh’.

¹⁰ See Berson (2021).

¹¹ See also Murdoch (2013, 27). For two recent discussions of the idea that ways of thinking are ‘ways of worldmaking’, as Nelson Goodman put it (1978), see Cappelen (2018) and Srinivasan (2019, manuscript).

¹² Delacroix (2022, 28, 78–79, 131) also develops the congenial notion of an ‘axiological habitat’ that we continually and collectively reshape.

¹³ See Steinbeck (2006, 150). It is the fact that concepts can take us beyond what we knew when we devised them that makes the image of stairs leading beyond themselves especially apt. In other words, it is possible to get more out of concepts than we put in.

hitherto hidden aspects and dimensions of the world, heighten our powers of perception, put previously inaccessible conclusions within our reach, cast things in a fresh light, unlock novel forms of reasoning, and open up new forms of knowledge. As a result, the concepts we construct can lead us beyond what we could have imagined when we constructed them, or prove valuable in ways we did not anticipate—that is the sense in which we might emerge ahead of our accomplishments, feeling that our concepts are smarter than we are.

Besides opening up new possibilities, moreover, concepts also *guide* us in choosing which of these possibilities to realize. Not only do they render certain objects and properties salient; they bestow on them a rational articulation and significance, thereby rendering us sensitive to certain *reasons* that can guide our thought and actions. As Jean Hampton observes, reasons possess a ‘guiding and commanding nature’ (1998a, 49). While she puts particular emphasis on the commanding nature of reasons, their guiding nature, which is a softer form of authority, is just as important.¹⁴ Reasons act, as often as not, through their ‘enticing’ rather than ‘peremptory’ force (Dancy 2004, 21). They commend as well as command, and condemn as well as commend. Which reasons we are commanded, guided, or enticed *by*, however, is a function of our conceptualizations—our concepts as we actually understand and use them.

This implies that the concepts we have at our disposal also *constrain* us. Already in highlighting and rendering salient certain aspects of the world, they obscure and blindside us to other aspects of the world. That reflects a structural constraint: to focus attention on an aspect of things necessarily involves detracting attention from other aspects—to pay attention to every conceivable aspect at once would be to pay attention to nothing. We can only make certain things visible at the price of making others invisible. Thought is inherently selective in this way, and to select ways of thinking is to select between different ways of being selective.

Downstream of the ways in which concepts channel and constrain our attention lie the ways in which concepts constrain what thoughts we can entertain and what desires and attitudes we can form. We make sense of things, and form attitudes towards them, under descriptions, and what descriptions are available to us is a function of our

¹⁴ For a related discussion of how the authority of reasons goes beyond the ‘authority of command’, see Laden (2012, ch. 3).

conceptual repertoire. The attitude of respect, for example, characteristically involves respect under a description, and I can only respect someone *as* a human being, or *as* a philosopher, if the concepts *human being* or *philosopher* are available to me.

Similarly, the conceptual architecture we inhabit constrains what we can intelligibly do. One cannot act *in the name of justice* if the concept of justice is unavailable; without a host of concepts in play, the best chess player is reduced to a literal wood-pusher; even something as simple as writing down one's name requires the concept of a *name*, and it takes an entire conceptual edifice for the same physical motions to count as the signing of an executive order. What meaning our actions can have thus depends notably on what concepts we can draw on in interpreting them. Thus, our concepts constrain us in that they delimit the horizon of what is possible for us. As Alexis Burgess and David Plunkett pithily put it, 'our conceptual repertoire determines not only what we can think and say but also, as a result, what we can do and who we can be' (2013a, 1091).

While the powers of conceptual architectures express themselves in a number of sometimes subtle forms, the most salient and distinctive way in which they hold sway over our lives is by sculpting the network of reasoning pathways we regard as correct. To think in terms of some concept *F*—*kitsch*, *sacrilegious*, *chaste*, *unpatriotic*—is not merely to discriminate between what is *F* and what is not-*F*, but also to treat the distinction as making a difference to what else one has reason to think or do. In adopting a concept *F*, one does not just come to recognize certain considerations as reasons for or against its applicability, but also comes to recognize its applicability as a reason for or against drawing certain further conclusions.¹⁵ In this way, the concepts we possess alert us to certain inferential relations between judgements: the applicability of one concept implies or excludes the applicability of other concepts—if something is a *villanelle*, it follows that it has nineteen lines and that it is not a sonnet.¹⁶ Concepts thereby sculpt the network of reasoning pathways we regard as correct.

¹⁵ This two-faced model of the articulation of concepts goes back to Michael Dummett's (1973, 434) generalization of Gerhard Gentzen's work on sentential connectives, and figures centrally in the inferentialist tradition (Brandom 1994, 2000, 2008; Dummett 1973; Koreň 2021; Kukla and Lance 2009; Peregrin 2014).

¹⁶ That is only one of several rigid formal requirements that a villanelle must meet according to the influential definition by the French poet Théodore de Banville. A well-known example is Dylan Thomas's 'Do not go gentle into that good night'.

We can acknowledge this formative power of concepts without endangering the objectivity of the reason relations in question. For even if the reason relations themselves are thought of as obtaining entirely independently of human concept use, it remains true that the concepts *we possess* shape the network of reasoning pathways *we regard as correct*. And the power to sculpt *that* network is a power so fundamental to so much else of philosophical concern that the authority question can seem, as Ernest Gellner remarks, like ‘the philosophically supreme question’.¹⁷

By sculpting the network of reasoning pathways we regard as correct, concepts shape and structure our lives. Consider how deeply the understanding and behaviour of a child is altered by acquiring the concept of sharing, or the concept of lying, or by learning that another’s suffering is a *reason* to try to help.¹⁸ Or consider how the trajectory of a human life can be altered through enculturation into the use of concepts such as *heroism*, *saintliness*, or *genius*; or how the character of a society varies with the availability of such concepts as *democracy*, *nationhood*, or *socialism*. Concepts can be recipes for flourishing, and Nietzsche was exaggerating only mildly when he remarked that a concept can also be a ‘formula for decadence’ (2005b, §15).¹⁹

We can thus think of the life-shaping influence of concepts as consisting chiefly, and most distinctively, in their power to license or exclude certain transitions in reasoning—not just transitions from thought to thought, but also from perception to thought and from thought to action. The concepts we use prescribe the patterns of thought we recognize as rational. This is true even if we accept that reason relations obtain between entire judgements rather than between the concepts in terms of which these are articulated, so that the primary form of conceptual content is propositional content. For even so, *which* reason relations we can appreciate at the level of propositions remains a function of which concepts we have at our disposal to articulate those propositions. Though concepts are not themselves reasons, they determine what we recognize as a reason for or against what by marking distinctions that in turn permit the articulation of claims that give us reasons for or against thinking or doing certain things.

¹⁷ See Gellner (1984, 260).

¹⁸ On that last example, see Wong (1991), who argues that the process by which one comes to think of suffering as giving one reasons to help is crucial to developing the virtue of compassion.

¹⁹ See Huddleston (2019) for a detailed exploration of this theme in Nietzsche.

Accordingly, adopting or abandoning a concept involves forging or severing inferential connections. We can expand our network of reasoning pathways by acquiring new concepts, and prune it by eschewing or losing concepts. An anodyne example is how introducing the concept *foul* into a game has the effect of licensing transitions in reasoning that were not licensed before, such as the move from ‘ x was tackled in a certain way’ to ‘the game should be interrupted’. Less anodyne examples are the concepts expressed by derogatory epithets that link the property of being a member of a certain ethnic group to some negative characteristic, such as being lazy, stupid, or cruel. People who think in terms of such an epithet can, by those conceptual lights, properly move from (1) ‘ x is a [insert ethnic group]’ via (2) ‘ x is a [insert derogatory epithet]’ to (3) ‘ x is [insert negative characteristic]’. For those whose conceptual repertoire does not include the relevant epithet, by contrast, there is no such direct inferential route from (1) to (3). The derogatory concept has built an inferential bridge that would not be there without it.²⁰

What is more, those who want to resist the transitions in reasoning made possible by these inferential bridges are forced to remove the bridge itself by rejecting the very concept expressed by the derogatory epithet. It is no use disputing that the conditions for the applicability of the epithet are met, since, by the crude light of concepts like these, anyone with the relevant ethnic characteristics straightforwardly meets them. Nor does it help to insist that, though the epithet applies, the attribution of the negative characteristic does not follow by default. It is the whole point of such concepts that it follows by default.

Of course, to say that it follows by default is not to say that it follows indefeasibly; reasoning by default can be thought of as involving defeasible generalizations.²¹ The reason relations encoded by our concepts are often non-monotonic: $F(x)$ may be a reason for $G(x)$ in the sense that if one accepts that $F(x)$, one has reason to conclude that $G(x)$ *by default*, but this reason may be defeated by the addition of further premises.

Moreover, the reason relations encoded by our concepts are one thing, the actual performances of human reasoning are another. As Gilbert Harman points out, the

²⁰ See Dummett (1973, 454) and Brandom (2000, 69–72) for an account of the epithet ‘boche’ along these lines. For a recent defence and elaboration of their account, see Mühlebach (2021, Forthcoming).

²¹ A default logic along these lines is developed by Harty (2012).

activity of reasoning remains in some respects underdetermined by the reason relations we accept as correct, for instance when relations of material implication fail to determine whether one should accept the consequent or reject the antecedent. ‘It is not always true that one may infer anything one sees to be implied by one’s beliefs’, Harman observes. ‘If an absurdity is implied, perhaps one should stop believing something one believes instead of accepting the absurdity’ (1984, 113). Ultimately, it is up to concept-users to decide when the needle on the compass of reasoning should swing from *modus ponens* to *modus tollens*.

It must also be granted that recognizing the authority of a concept and its concomitant reasons is not the same as actually *acting on* the reasons it adverts to in the particular case: I might generally recognize the authority of the concept of honesty, but the temptation to lie to my advantage might carry the day this time around. Many concepts plausibly emerged precisely to stake claims against the all too real temptations not to behave as these concepts give us reason to. Ethical concepts have been thought to stake claims against self-interest,²² for example, amplifying and redirecting altruistic tendencies that would otherwise remain overly partial, in both senses of the term.²³ More generally, new value concepts rendering us sensitive to new reasons and motives can remediate the shortcomings of the motives we have anyway already.

Consequently, moral progress happens notably through alterations to our conceptual repertoire: if one’s behaviour is a function of the reasons and motivations one is sensitive to, it can be improved by acquiring new concepts and their concomitant reasons and motives. When it comes to ‘morals and politics’, John Stuart Mill notes in his *Autobiography*, ‘improvements in the lot of mankind’ require changes in the ‘constitution of their modes of thought’ (1874, 239).²⁴

Yet morally significant differences need not always be reducible to behavioural differences in the public forum of action and choice. When Murdoch urges that ‘we need more concepts in terms of which to picture the substance of our being’, because ‘it is through an enriching and deepening of concepts that moral progress takes place’ (1961,

²² See Williams (1973b, 250).

²³ See Kitcher (2011, 86)

²⁴ On the historical connection between political innovation and conceptual change, see Ball, Farr, and Hanson (1989).

20), she is not thinking merely of getting people to *behave* better. Her point is that morally significant differences can consist simply in how, *in foro interno*, one conceptualizes a situation:²⁵ ‘Moral differences can be differences of concept as well as differences of choice. … How we see and describe the world is morals too’ (1999, 73). Williams echoes her point when he stresses that ‘an extremely important form of ethical difference’ is ‘that between those who do and those who don’t use a certain concept’ (1995k, 237). If this is right, it suggests that moral progress can be not just facilitated, but constituted by alterations to our conceptual repertoire. Adopting or abandoning concepts can be progress in itself.

1.3 Expressing the Authority Question

In light of these various powers exerted by concepts, it makes sense to ask what reasons one has to abide by a given concept—why, as Dworkin says, one should accept it. The great benefit of being able to raise the authority question and think about second-order reasons is that it enables us to assess and revise our conceptual architecture from within. Yet we rarely question our use of a concept in quite so explicitly metaconceptual terms. What linguistic forms does the authority question actually take?

The authority question can find linguistic expression in a variety of more idiomatic locutions. One person might try to dissuade another person from doing *x* by urging that *x* is *hypocritical*, for example, to which the other person might react by asking for reasons to think that *x* is hypocritical, or by querying, more radically: ‘So what? Why should I care?’ Explicated in terms of the present account, this amounts to asking: ‘So what if that concept yields this conclusion? Why should I accept that concept to begin with?’ This is to demand a reason for a reason—a reason to recognize the authority of the concept of hypocrisy and the reason it gives those who recognize it.

Other natural locutions allow one to express the same demand. When someone commends an action or a substance for being ‘natural’, for instance, they might encounter the reply: ‘Why is that a reason?’ Or when someone warns that video games cause harm, the discussion might soon revolve around what ‘counts as’ harm. Or when someone

²⁵ See Murdoch (1956, 1961). On the moral significance of articulating new moral concepts to reduce the dissonance between ethical experience and received conceptual resources, see Congdon (2023).

justifies a claim about a person by saying ‘Because she’s a woman’, the question might come back: ‘How does that follow?’ In each case, one speaker relies on a conceptualization whose authority the other speaker challenges.

There are thus various well-established ways of raising the authority question without expressing it explicitly in terms of whether one has reasons to use a certain concept. One might question the authority of a certain concept by asking why something’s being *F* is a *reason* for something else, or what *counts as F*, or what *follows from* something’s being *F*. These are well entrenched expressions that enable us to question and adjust our conceptual repertoire. Natural languages already come equipped with metaconceptual expressions that allow us to question the authority of concepts, and thereby to criticize and revise the network of reasoning pathways we regard as correct.²⁶

At the limit, one might even question the authority of a concept through the right combination of ‘Why?’-questions with locutions that bluntly express the dismissal of the rational authority of some consideration, such as ‘So what?’ To see this, consider first the question of what conceptual resources are required simply to think that some consideration is a *first-order* reason for another consideration. As Andreas Müller remarks, that is one of our most fundamental agential capacities, displayed already by young children, and it cannot be all that demanding:

Anyone who is able to answer requests for justification such as ‘Why do you think that p?’ or ‘Why did you do A?’ thinks that something counts in favour of believing that *p* or doing *A*. They do so even if they cannot express this by using the term ‘reason’, or indeed in any other way than by answering that question. It thus takes very little to prove that you have the conceptual resources to think that some considerations are reasons for (or count in favour of) certain beliefs or actions: being competent with the justificatory use of ‘because’ in answering such ‘why’ questions is sufficient. (Müller 2019, 6)

Similarly, anyone who is able to react to the giving of some first-order reason such as ‘Because it’s *natural*’ with ‘So what?’ is capable of expressing their willingness to question

²⁶ Here I draw on the Carnapian idea, arguably adumbrated by Wittgenstein and Ryle, that alongside our empirical, ground-level vocabulary performing descriptive functions, natural languages include metalinguistic vocabulary with the function of explicating and regulating the norms governing our ground-level vocabulary. The idea was elaborated notably by Sellars (1958), and has more recently figured prominently in the work of Heal (2007), Brandom (2015b), and Thomasson (2020a).

the authority of the concept in this connection—some more back and forth may be required to determine that this is indeed what they are doing, but they certainly need not resort to high-flown terms like ‘second-order reason’ or ‘authority of concepts’ to express their doubts over whether being ‘natural’ should count as a reason here. Some degree of conceptual competence with the notion of a second-order reason is far more widespread than the technical language whereby we can make it explicit.

In fact, as David Plunkett and Tim Sundell have persuasively argued, many ordinary disputes that are ostensibly about object-level issues are actually better understood as metaconceptual disputes, i.e. disputes about which concepts to use.²⁷ Besides genuine disagreements between parties who share a concept and disagree about whether something falls under it and merely apparent disagreements between parties who talk past each other because they express completely different concepts by the words they use, there are also interesting cross-over cases, where parties genuinely disagree *and* take a word to express different concepts. In those ‘metalinguistic negotiations’, as they call them, the parties to the negotiation each use (rather than mention) the concept that they are thereby advocating as preferable to alternatives. We commonly do this when introducing someone to a concept, saying things like: ‘Sebastian is *debonair*, for instance’. When someone else demurs: ‘No way, Sebastian is not *debonair*!’, the disagreement need not be about Sebastian—it can be about how the concept *debonair* should be understood. In this case, the concepts being advocated are used to show what kinds of judgements they each license and exclude. And though the sentences are not ostensibly about concepts, they are being *put to metaconceptual use*, allowing the two speakers to negotiate how best to define the concept at issue. As Plunkett and Sundell emphasize, such metaconceptual disputes conducted in object-level terms are a regular fixture in colloquial discourse.

The fundamental insight here is that since object-level judgements express and exemplify concepts, we can also use them to debate which concepts to use. If, as Wittgenstein asserts, agreement in concepts requires agreement in judgements, the contrapositive of that thought is that disagreement in judgements can express disagreement in concepts, which is to say disagreement over which concepts to use. This

²⁷ See Plunkett (2015) as well as Plunkett and Sundell (2013a, b; 2019) and McPherson and Plunkett (2020); see also Stroud (2019).

connection between concepts and the substantive object-level judgements we make with them is what grounds the insight that some ostensibly first-order disagreements are really metaconceptual disagreements.

Once we recognize that we can debate which concepts to use *by* using them, however, it is only a small step to the further observation that we can debate not just whether to adopt or abstain from a concept altogether, but which concepts to use in a particular situation. As Sarah Stroud has pointed out, in debating whether the main takeaway from a shared experience should be that x is F or that x is G , we are not contradicting each other, but disagreeing about what kind of conceptualization the situation calls for: we are disagreeing about which aspects of a situation are most worth fastening on, and wherein their significance lies.²⁸ As Stroud argues, this is also, broadly speaking, a question in conceptual ethics. It is merely that it tackles the question of which concepts to use from its most applied and particularized end, asking, in the first instance, not which concepts to use *tout court*, but *when* to use which concepts. This may not be as self-consciously metaconceptual, or as radical, as explicitly raising the question of whether to go in for any kinds of judgements in terms of F at all—eschewing a concept in a particular situation is, after all, not yet to eschew it altogether. But in taking a stand on what merits notice in a given situation, we are also implicitly endorsing certain concepts over others, and encouraging certain conclusions over others—we are voting with our thoughts.

What is more, which way we vote tends to affect the outcome, which is to say that the concepts we foreground in grasping a situation affect the choices we make on that basis. This comes out most clearly when a difficult choice turns on how one conceptualizes things. The difficulty of a tough choice need not reduce to the difficulty of choosing between different options given how one conceptualizes them. It can lie, at least in part, in the difficulty of choosing between different ways of conceptualizing each option, especially when one knows that these carry different implications and that one's choice of conceptualization therefore affects one's attitude towards the resulting options. That is why people often debate how to conceptually frame what they do ('This is F ', one exclaims; 'I prefer to think of it as G ', the other retorts).

Consider José Bermúdez's (2021, 7–8) illustration of the importance of framing: when Aeschylus describes Agamemnon's anguished choice between sacrificing his

²⁸ Stroud (2019, 25).

daughter to comply with the will of the goddess Artemis or renouncing his conquest of Troy, Agamemnon's deliberations do not just involve weighing two painful outcomes against each other while holding their characterizations fixed; different ways of conceptualizing the same action compete in his mind: should he think of the action the goddess demands of him as *pious obedience*, or should he think of it as *slaying his child*? As *doing what is right and holy* or as *shedding virgin's blood*? The chorus describes his agonized train of thought as setting out from, and eventually circling back to, notions of obedience and lawfulness ($\theta\acute{\epsilon}\mu\iota\varsigma$).²⁹ It is not just in his eventual action, but already in foregrounding these conceptualizations that Agamemnon submits to the will of the goddess.

1.4 When the Authority Question Arises

What prompts the authority question in the particular case? 'Paper' or 'tin' doubts cannot motivate inquiry, C. S. Peirce remarked, and, as Wittgenstein added, doubt is no less in need of reasons than confidence.³⁰ The force of the authority question is thinned out if it is applied evenly across our entire conceptual repertoire at once. It is most forceful when focused on particular concepts in particular situations that specially invite the question. So what might raise the authority question?

A clue is provided by the fact that Dworkin presents the question specifically as a challenge for composers of definitions. This suggests that when people compose new definitions, they incur the burden of showing why one should accept those definitions. And while composing definitions is by no means restricted to philosophy, it is clearly central to what philosophers do. Equally central to philosophy, therefore, is the task of demonstrating the authority of the concepts that come out of this activity. Nietzsche conceives of post-Socratic philosophy in exactly these terms: 'What Plato and fundamentally all the post-Socratics did: that was a certain legislation of *concepts*' (85:34[84]).³¹

Calling it 'legislation' expresses Dworkin's intended meaning better than Dworkin's

²⁹ See Aeschylus (1966, l. 205–17). On the difficulties of translation raised by this passage, see Nussbaum (1988, 35, 431 n. 36), but the point holds, *mutatis mutandis*, for different translations as well.

³⁰ See Peirce (1931, 5.416) and Wittgenstein (1969, §§4, 122, 323, 458, 519).

³¹ Charles L. Stevenson also underscores Plato's proclivity for 'persuasive definitions' (Stevenson 1938; Stevenson 1944).

own ‘composition’. It re-describes what might otherwise sound like an individualistic artistic activity in explicitly social and normative terms, thereby underscoring how holding out concepts for others to accept entrains a correlative responsibility to demonstrate, *to* the people who are urged to adopt the concepts, why they have reason to adopt those concepts and structure their affairs in terms of them.

The authority question thus naturally arises when a concept is explicitly announced as revisionary, as it typically is when philosophers—or theorists in other disciplines—present themselves as conceptual innovators or engineers rather than as analysts of concepts. While analysis can still pretend to be to concepts what ornithology is to birds, engineering openly tampers with our concepts. Because of this, engineers tend to offer an explicit rationale for their revisionary intervention. A tempting idea in this context is that of the philosopher as a conceptual plumber—as someone who *fixes defects* in our concepts.³² The engineered concept is then offered as authoritative insofar as and because it is free of some defect inherent in the concept it is meant to replace, such as its vagueness or its tension with some other concept. Once the supposedly improved concept begins to overturn the judgements encouraged by our entrenched ways of thinking, however, the authority question re-emerges in relation to the supposed improvements: why should we care so much about tidying up our concepts in certain ways, if the price is to overturn judgements that more immediately have force with us? This will be the topic of Chapter 4.

The task of demonstrating the authority of the concepts is not just a task for conceptual engineers who aim to prescribe new concepts, however. It is a task faced also by conceptual analysts who aim to describe the concepts we use already. When someone puts forward a particular analysis of a concept for us to adopt, this also invites, if perhaps less obviously, the question of why we should accept that analysis.

Imagine a theorist who takes a contested and somewhat indeterminate concept such as *democracy* and presents you with a weighty tome entitled *Democracy: An Analysis of the Concept*, which articulates and draws out the ramifications of a very definite and precise concept of democracy. Suppose that, on this concept, it clearly follows that your supposedly democratic government is in fact not truly democratic. On that particular

³² For the idea of philosophy as conceptual plumbing, see Midgley (1996). For the idea of conceptual engineering as a matter of fixing defects, see Cappelen (2018).

concept of democracy, this may well follow—but why, you will naturally want to ask, should I think in terms of *that* concept? What is there to show that this particular conceptualization of democracy is more than a scholarly idiosyncrasy, or worse, a form of conceptual gerrymandering designed to validate the political preferences of the analyst?³³

Appeals to intuitions hardly help, as these can be equally idiosyncratic or ideology-driven.³⁴ The clearer and the more determinate the proffered conceptualization of democracy, the more pressing the question of what reasons we have for endorsing this particular pattern of reasoning. Even if we are prepared to accept the proposed concept on the grounds that it exhibits theoretical virtues such as clarity, determinacy, precision, and consistency—an influential idea we shall return to in Chapter 4—the authority question sharply reappears when many such proposals compete, and we are forced to adjudicate between them. Some basis is required for distinguishing between authoritative definitions and mere gerrymandering with concepts.

The line between analysis and engineering is itself blurrier and shiftier than this simple delineation suggests, moreover. Conceptual analysis is frequently taken to encompass not just the empirical description of concept use in a community, but also the clarification of how a given concept *should* be understood for the purposes of a certain line of argument or a research programme.³⁵ And the line between analysis and engineering can also be blurred simply by changes in perspective or context. The most visionary attempts at conceptual engineering may, by dint of their own success, acquire the marmoreal dignity of classic analyses—Plato, Descartes, Hobbes, Kant, or Freud can seem to us now merely to describe deeply entrenched concepts that they in fact helped create. Or consider the conservative analyst who, in rapidly changing circumstances, insists on extending the life of a received concept; he may invite more questions regarding the authority of his analysis than the engineer who understands which way the wind

³³ See Baz (2017, 46–48) for an articulation of this worry, which he terms ‘cognitive diversity skepticism’. Related worries are highlighted by Eklund (2017, 13), Clarke-Doane (2020, 180), and Wolf (2020), who sketches a Hegelian answer to it.

³⁴ See Machery (2017, ch. 4) for a critique of conceptual analysis on these grounds.

³⁵ For accounts of conceptual analysis which emphasize its breadth and richness as well as the ways in which it differs from the empirical description of concept use as practised by linguists and anthropologists, see Glock (2017) and Rathgeb (2020).

blows and revises the concept to move with the times.³⁶ In addition to these extrinsic developments that may give engineering the authority of analysis or analysis the contentiousness of engineering, there are also pressures *within* analysis and engineering that can drive the one to shade into the other. Analysts' quest for differentiation and determinacy may easily lead them beyond established use, while engineers may find their proposals more readily taken up when proffered under the mantle of analysis.

Besides cases in which someone puts forward a concept for us to adopt, the authority question can also become acute when we lose confidence in the concepts we use already. Among the experiences that can prompt the authority question in this retroactive form is the experience of conflict between concepts. Different concepts of course frequently recommend starkly different courses of action. But in some cases, equally deeply entrenched concepts systematically pull in different directions and are impossible to realize in concert, not just because circumstances are unpropitious, but because the concepts themselves inherently and irremediably conflict. This makes it tempting to try and dissolve the conflicts by questioning the authority of one of the concepts involved and finding some reason to discount and ignore the claims it makes on us—either by bracketing them just this once, or by permanently disqualifying them from weighing into our deliberations. Utilitarians, for example, find reasons for reasons in the fact that the use of certain concepts tends to contribute to overall welfare, and, on that basis, resolve conflicts by making conceptual authority a function of utility.

It is worth emphasizing that the discomfort of conceptual conflict is not the preserve of theorists with a taste for tidy systems. The felt necessity to alleviate or dissolve tensions between concepts has deeper roots in the fact that living by concepts which systematically pull in divergent directions can be deeply uncomfortable, because it presents one with choices that cannot be resolved without loss. This is a widely shared experience that may prompt the authority question in anyone, and may indeed underlie the craving for tidy systems. Accordingly, the aim of relieving conceptual tensions has been thought by some to be justified by a practical concern with well-being: on Valerie Tiberius's (2018) value fulfilment theory, for example, having conflicting concepts counts as a form of 'ill-being', and the pursuit of well-being centrally involves alleviating

³⁶ By the same token, concept preservation may warrant being described as a form of engineering when moving with the times would amount to backsliding; see Lindauer (2020).

or dissolving conflicts within and between our value concepts. A salient way to achieve this is by questioning the authority of the concepts involved.

Another kind of situation which may prompt the authority question is when we take concepts out of their usual context of application and project them into a radically novel context. The novel context can be a product of technological advances—the metaverse may not call for the same set of concepts as the universe. But such projections from a source context into a new target context also arise when a concept is transplanted from one academic discipline into another, or when concepts are extended from one sphere of application to a very differently configured one. It is open to question, for example, to what extent the concepts of human psychology can help us make sense of animal minds, or to what extent the concepts of domestic politics can be made to fit international politics.

What then prompts the authority question is the sense that the novel context is not obviously covered by our existing concepts and remains conceptually underdetermined, which renders acute the question of how to conceptualize it. Where inquiry is as yet insufficiently *framed*, as Céline Henne (2022) argues, we must engage in a different sort of inquiry, *framing inquiry*, which creatively extends our conceptual repertoire to cover novel contexts. But just because this sort of inquiry is self-consciously involved in framing the previously unframed, it invites scrutiny of the authority of the proposed conceptualizations. Perhaps animal minds are different enough to call for their own set of conceptualizations, or perhaps the peculiar configuration of international institutions requires us to rethink the way we conceive of legitimacy, democracy, representation, and other concepts in the international sphere.³⁷ In each of these examples, the question arises: should we be thinking in those terms at all here? This is only to raise the authority question in a domain-restricted form, however: it does not question the authority of the concept across all domains, but only its authority within the novel domain.

Finally, we are sometimes led to question the authority of particular concepts for reasons that fall out of the concept itself. Take Sabina Lovibond's description of how a user of the concept *first-class mind* might come to wonder whether she is guilty of conniving participation in the use of a problematic notion. Is the concept perhaps

³⁷ Such reconceptualizations have been advocated, along different lines, by Krisch (2010), Cohen (2012), and Cueni (2020), for example.

vacuous, like *phlogiston*? Is it being used merely as a pretext for differential treatment, like *witch*? As Lovibond observes, one may be on the fence about this: ‘Are there such things as “first-class minds”? A frank answer might be: sometimes I talk (and think) as if I believed in these things and sometimes not’ (2015, 137).³⁸

Anxiety that a concept might turn out to be vacuous, i.e. have an empty extension, is not the only reservation one might have about a concept. Take Oscar Wilde, whose then-notorious dictum that there was no such thing as an immoral book was put to the test during his trial: the prosecutor challenged Wilde to deny that a story circulating among Oxford undergraduates at the time (and falsely attributed to Wilde) was *blasphemous*. Wilde was willing to grant that it was ‘disgusting’ and ‘horrible’, but, despite being repeatedly pressed to do so by the prosecutor, Wilde refused to enter into whether it was blasphemous or not. By any reasonable definition of *blasphemous*, the story, entitled ‘The Priest and the Acolyte’, fell under the concept—the concept was certainly not vacuous. Instead of denying that the concept applied to story, therefore, Wilde declared: ‘“Blasphemous” is not a word of mine.’³⁹ This was not a denial of anything the prosecutor asserted. It was a refusal even to think in these terms.

Another example of discomfort about concepts already in use is offered by Christine Korsgaard, who discusses the concepts *masculinity* and *femininity*, and points out that, for critics of these concepts, they are ‘straitjackets, stunting everybody’s growth’ (1996, 77).⁴⁰ The right response to someone who reproachingly points out that a certain behaviour is not masculine/feminine, therefore, is not to retort that this claim is false, or to insist that the behaviour is in fact masculine/feminine; the right response, Korsgaard writes, is: ‘let’s not talk that way’ (1996, 77)—by which she means, presumably, not just that we should stop talking in those terms, but that we should cease even to *think* those terms, and hence cease to hold ourselves and others to the ideals enshrined in those concepts.

³⁸ See Lovibond (2015, 136–138). While she also describes the thinking process of someone who reaches a position of outright disbelief in first-class minds, her interest is in the concept-user who remains agnostic on the issue.

³⁹ See Hyde (1973, 107). The story in question was in fact written by John Francis Bloxam, editor of *The Chameleon*, the magazine in which the story appeared.

⁴⁰ See also Wollstonecraft (2014, 30–33, 48, 158, 204) and Mill (1988, IV, 86–88).

Lovibond, Wilde, and Korsgaard all single out particular thick normative concepts and question their authority.⁴¹ They gesture towards a lack of reasons to use the concepts in question, or even towards reasons *not* to use them. In the absence of a theory of concept use, however, it is hard to specify the nature and range of these reasons, and how one might more systematically go about answering the authority question.

What is clear, however, is that there are many situations that can prompt us to raise the authority question—and not just in order to evaluate different proposals for conceptual innovation, but also to reevaluate the concepts we were using already. The demand for reasons that rationalize the recognition of certain concepts and their correlative reason relations as authoritative arises for concepts old and new.

1.5 The Autoethnographic Stance

Assessing the authority of a concept requires one to relate to it in a quite peculiar and philosophically interesting way. The mere ability to *mention* a concept ('They propose this concept of "F", whatever that is') is not by itself sufficient. Some grasp is required of how things present themselves to one who actually *uses* the concept. One must understand the concept 'from the inside' in order to assess its claim to authority.

At the same time, one must find some way of stepping back from the concept sufficiently to acquire critical leverage over it—it would be too uncritically accepting of the concept simply to insist that we should use it because there is so much *F*-ness around and *F*-ness is important for the reasons the concept *F* itself advertises to.

To make sense of what is involved in engaging with the authority question, we therefore have to distinguish not just between *mentioning* and *using* a concept, but also between two different ways of *using* a concept.⁴² Adrian Moore (2006b, 137) has usefully marked this distinction in terms of the contrast between *disengaged* and *engaged* concept

⁴¹ For other examples of suspect concepts and discussions of the theme of concept loss, see also Murdoch (1961), Diamond (1988, 2021), Teichmann (2021), and Mulhall (2021, 28–31).

⁴² The use/mention distinction alone, if it is to be clear-cut, cannot capture the difference at issue here, because, as Moore (2019b, 15) argues, that difference turns on what a concept is employed *for*, and, on any account of the use/mention distinction that renders it clear-cut, the mere fact that a concept is mentioned cannot tell us much about what the concept is employed for.

use.⁴³ This distinction between engaged and disengaged use is crucial to understanding what is involved in addressing the authority question.

When using some concept *F* in an engaged way—the concept *funny*, say—one looks through it at the world, as if through a lens, and is immediately responsive to the gestalt it gives the world, the aspects it renders salient, and the reasons it advertises. It is then perfectly correct to describe the world as containing plenty of things that are *F*, and to insist that we use the concept because there is so much *F*-ness around. An essential part of what is involved in grasping the concept, moreover, may be to conceive of what it picks out—*F*-ness—as being entirely independent of human observers and their practices of concept use. If a dinosaur once stumbled in a particularly funny way, this would still have been, as a matter of fact, objectively funny, even if there were no human observers around then who could have perceived it as such.

At the same time, we can also think *about*, rather than *with*, the concept *F*, standing back from it to hold it up to reflective scrutiny. This is what we must do to critically assess a concept's authority. We must take a sideways look *at* the lens instead of peering through it. But to understand what we are looking at, we still have to grasp the concept, and grasp what the world looks like to one who peers through it. To think about a concept in this comprehending way, we thus still have to think *with* it in an important sense. We must use the concept in a disengaged way in order to be in a good position to evaluate whether to use it in an engaged way.

Thick normative concepts offer a particularly clear illustration of this: they neither just describe nor just evaluatively label independently describable patches of the world. Rather, what forms part of their extension is itself a function of the evaluative attitudes of those who deploy these concepts, and therefore their extension cannot be neutrally specified, without adopting an evaluative stance—the extension is *shapeless* without the evaluation.⁴⁴ As T. M. Scanlon unpacks the point: ‘In order to trace the contours of the

⁴³ Which itself crystallizes a distinction prefigured in Williams (1986, 203–4; 1995i, 207; 2002, 50–51; 2011, 157). See also Thomas (2006, 146) and Goldie (2009) for other illuminating ways of characterizing the distinction.

⁴⁴ See Dancy (1995), Kirchin (2010), and Roberts (2011, 2013) for that way of putting the point. For articulations of the same point in more general terms, see Williams (1995a, 563; 1995i, 206; 1996, 29; 2011, 157), McDowell (1998a, b), Scanlon (2003, 276), and Anderson (2004, 14).

ethical concept's applicability we have to understand its evaluative point ... we must be guided by the evaluative perspective of a thick concept in order to apply it' (2003, 276). To think intelligently about a thick concept, one must take up, at least imaginatively, the evaluative perspective of those who use the concept in an engaged way.

A model for this way of relating to a concept can be found in the way an ethnographer studies a concept from another culture. An ethnographer can become an expert on some mystical concept of another culture without having to live by the concept herself: she need not structure her own affairs in those terms and may remain unresponsive to the distinctive reasons the concept articulates.

This is to adopt what Williams calls 'the ethnographic stance', a stance from which one 'has an imaginative understanding of a society's ethical concepts and can understand its life from the inside, but does not share those concepts' (1986, 203–4)—one can 'pick up and understand' those concepts, grasping how they are integrally related to other things of value to those who use them, and how they engage their emotions, 'and yet reject those concepts' (2021, 278).

We already adopt this stance towards the concepts of religions we have some familiarity with but do not practice ourselves. Moore offers the example of the concept *Sabbath*.⁴⁵ Disengaged use of the concept is sympathetic enough to enable one to grasp the concept's role and place in other people's lives—'sympathetic' in the sense of feeling one's way into another perspective to the point of resonating with its sentiments and concerns (much as the sitar's 'sympathetic strings', though untouched, resonate with the primary strings).⁴⁶ But the disengaged concept-user grasps all this merely in the spirit of an ethnographer, without living by the concept herself: she does not structure her own affairs in those terms—she does not *observe* the Sabbath, as we naturally put it in this case. Engaged use of the concept, by contrast, involves not just sympathy, but identification: it involves living by the concept, i.e. being emotionally and rationally responsive to it and its concomitant reasons in the conduct of one's own affairs.

⁴⁵ See Moore (2006b, 137).

⁴⁶ For the sitar analogy, see Baillie (2000, 52). Hume aptly describes sympathy as the capacity of the 'minds of men' to be 'mirrors to one another' (Hume 2000, 2.2.5.21), and on his account, as on Adam Smith's, sympathy forms the foundation of impartial concern for general welfare—though see Sagar (2017) for the subtle differences in how Hume and Smith understand sympathy.

Disengaged use is thus sympathetic enough to grasp the concept from the inside, but not identified, whereas engaged use is both sympathetic and identified.

The human ability to adopt this ethnographic stance is an underappreciated but crucial influence on philosophical reflection. By enabling us to genuinely understand ethical concepts other than our own, it awakens us to the enormous variety to be found across human ethical thought, and hence to the realization that there are plenty of alternatives to many of our own concepts. Were we permanently locked in the engaged perspective, we could never ascend to the vantage point from which questions of objectivity, tolerance, and relativism become salient. That, presumably, is why Williams considers the fact that the ethnographic stance is possible to be important for moral philosophy.⁴⁷

An equally important point left unaddressed by Williams, however, is that we are also capable of adopting something akin to the ethnographic stance towards *our own* concepts—we can take up what might be called the *autoethnographic stance*.

The autoethnographic stance enables us to make disengaged use of concepts we normally use in an engaged way. This is not the same as viewing our concepts the way people living by different concepts might view them; it is still to view them as our own concepts, but against the backdrop of the fact that they are one set of concepts among a range of alternatives. Clifford Geertz eloquently describes the difference:

To see ourselves as others see us can be eye-opening. ... But it is from the far more difficult achievement of seeing ourselves amongst others, as a local example of the forms human life has locally taken, a case among cases, a world among worlds, that the largeness of mind, without which objectivity is self-congratulation and tolerance a sham, comes. (1983, 16)

Unlike the ethnographic stance, the autoethnographic stance first requires us to disengage ourselves from at least one of our concepts: we must cease to be fully absorbed in its use and its concomitant view of the world, and cease to be immediately responsive to its correlative reasons; but we also have to remain able to understand the concept from the inside and retain a detached, ‘offline’ responsiveness to its distinctive reasons, because

⁴⁷ As Williams avers in response to Simon Blackburn’s review of *Ethics and the Limits of Philosophy*: ‘The fact that the ethnographic stance is possible seems to me very important for moral philosophy’ (1986, 204).

serious reflection on the concept's merits requires us not merely to mention the concept, but to really use it, albeit in a disengaged way that embeds the use of the concept, and the perspective from which it is used, within a wider perspective from which that perspective and the way of thinking it enables can be appraised. We must disengage ourselves from the concept while stopping short of losing the concept or becoming estranged from it to the point of incomprehension.

When adopting the autoethnographic stance, we thus hover in a mid-level position, halfway between abandonment *to* and abandonment *of* the concept. This allows us to reflect on why we use it, and what makes it a good or bad concept for us to use. We cannot do this with all our concepts at once, for we still need to retain some conceptual basis *from* which to think about the concept under scrutiny. But we can adopt the autoethnographic stance in a piecemeal fashion, disengaging ourselves from one concept, or one connected set of concepts, after the other, and relying on the rest of our concepts to embed it into an evaluative perspective from which it can be evaluated. Throughout this process, we continue to be able to grasp what the world looks like to one who lives by the concept—what gestalt it gives the world and what reasons it advertises—but have temporarily suspended its correlative reasons to weigh our reasons for thinking in terms of such reasons.

The possibility of occupying this mid-level position, where one ceases to be fully absorbed in the use of a concept while remaining able to use it in one's reflection on its merits, is the basic fact from which this book sets out. It is only thanks to this remarkable ability to take up the autoethnographic stance that we can look for *reasons* to reason in terms of our concepts. Instead of being forced either to use or not to use certain concepts, the autoethnographic stance allows us to use them in a disengaged way, in order to assess their claim to authority over our lives.

Distinguishing between engaged and disengaged use also allows us to distinguish two corresponding modes in which one can raise the authority question about thick normative concepts. While I have emphasized that the possibility of using concepts in a disengaged way is key to raising the authority question in the mode at issue here, one can also raise it about concept *F* while using concept *F* in an *engaged* way, looking *through* it rather than sideways at it. This means that, for any thick normative concept, we can give reasons to use it according to the following schema:

The most compelling reason for us to use the concept *F* is that, given our concerns, our capacities and limitations, and the circumstances in which we live, we need to be suitably sensitive to the presence of *F*.

On this schema, *F* is thought of as being there anyway already, waiting for its presence to be picked up on by concept-users. The reference to our concerns, capacities and limitations, and circumstances will be explicated and further motivated in the chapters to follow; but for now, its point is to acknowledge that even if we derive the reason for our use of the concept *F* from the presence of *F*, something still needs to be said to account for our *interest* in the presence of *F*. *F* does not automatically inscribe itself into our conceptual apparatus just because it exists, and *a fortiori* not necessarily under that description. That is why, even if we picture the world as teeming with thick normative properties, some human concerns must be appealed to in order to motivate conceptual sensitivity to these properties.

But the problem with this kind of answer to the authority question, given from an engaged perspective on the concept in question, is not so much that it is wrong as that it is *too internal* to one's use of the concept. It makes answering the authority question too easy. From the engaged perspective of someone who thinks with or through the concept *F*, the world appears full of *F*, and it seems nearly inevitable that human beings would eventually have reason to come to think in terms of *F*. The authority of *any* concept with a non-empty extension can be vindicated following this schema.

Yet it is nonetheless important to acknowledge that the question can be addressed in this engaged mode, because this enables us to find a place for, and account for the plausibility of, the kinds of reason statements that refer either to the object of the concept in question or to the reasons the concept itself opens our eyes to. But tackling the question in this mode offers us too little critical leverage and does little to help us discriminate between more and less authoritative concepts.

If the authority question is easily answered by the engaged concept-user, for whom the salient ubiquity of *F* virtually compels us to think in terms of *F*, the question is a far more open one to the disengaged concept-user, who wonders how much reason to use the concept *F* we still have if we characterize our situation *without* drawing on that concept. From this perspective, the thought that we use the concept *F* because there is so much *F* around will hardly provide a *compelling* reason to use the concept, since this

perspective makes one too conscious of the fact that the impression that there is so much *F* around is partly a *product* of our use of the concept *F*, and therefore hardly well suited to provide an *independent reason* for it. The question, from this perspective, will rather be whether there are also reasons for us to use this concept that do not themselves draw on the concept, and instead adhere to the following schema:

The most compelling reason for us to use the concept *F* is that, given our concerns, our capacities and limitations, and our circumstances as characterized without drawing on the concept *F*, we need to use the concept *F*.

If reasons can be found that fit this schema, this will not merely be the concept patting itself on the back, but an independent vindication of its authority. That is why the possibility of taking up the autoethnographic stance is crucial to raising the authority question in the way that matters.

1.6 Conflating Engaged and Disengaged Use

But the possibility of taking up the autoethnographic stance brings risks as well as opportunities. Just because it opens up a second perspective on our concepts, it entrains the risk of conflating the two perspectives.

From the disengaged perspective, we may well conclude that the concept of funniness, for example, is not a concept that creatures differing sharply from us—such as dinosaurs—would have shared. Though there can be little doubt that many extremely funny dinosaur misadventures must have taken place over the course of the roughly one hundred and fifty million years during which they dominated the earth, all that funniness is likely to have been lost on them.⁴⁸ *Funny* is probably a distinctively human concept, not merely in the sense of being *possessed* only by humans, but in the sense of being *expressive* of our humanity.

This makes it tempting to infer that any funny misadventures that occurred before humans came on the scene were not, in fact, *objectively* funny. Indeed, having recognized, from the disengaged perspective, that the concept is a distinctively human concept, it is

⁴⁸ The example is loosely based on Williams (2014a, 380); see also his discussion of distinctively human values in Williams (2006c).

tempting to go further and conclude that, even from the engaged perspective, the right thing to say is that things are only ever ‘funny for us’.

That pressure to relativize the validity of our concepts is even stronger in the case of thick normative concepts that exhibit a great deal of variety *within* the range of human arrangements. Consciousness of this variety can exert pressure on our confidence in our own concepts—and, by extension, on the judgements we form with them. ‘What truth is that’, Montaigne exclaimed, ‘which these mountains bound, and is a lie to the world beyond?’⁴⁹ Fuelled by an awareness of alternatives, a nagging sense that we *could* think differently hovers over our dealings. This makes it almost irresistible to conclude that what *we*, in the contrastive sense that opposes some more or less narrow *us* to other people, perceive as reasons, truths, and facts, are really only ever reasons, truths, and facts *for us*, in virtue of the concepts we use, but not for other people.

This hovering ‘could’, this awareness of alternatives, can invite a relativistic reinterpretation of our own thoughts which it is important to resist. This relativistic reinterpretation is made all the more tempting by the fact that there really are some concepts that are properly understood as relativizing their own claims to applicability and validity—in matters of etiquette, for example, which are sometimes correctly understood to exempt outsiders, or to be applicable only in certain locations; or in the case of concepts such as *cocktail party*, where what counts as one is correctly understood to depend on certain people’s disposition to treat it as one.⁵⁰

Generalize this model, however, would result in a misunderstanding of most of our concepts. ‘When in Rome, do as the Romans do’ may be true enough for certain norms of etiquette, but it does not generalize to other norms.⁵¹ The thought that our reasons are only simply there *for us* pertains to the disengaged perspective, and to try to introduce it into the engaged perspective is to create a kind of chimera. The pupil who treats the sentence ‘2+2=4 for us’ as a sentence in arithmetic has misunderstood something. One has not mastered these mathematical concepts unless one recognizes their claim to universal applicability and validity. The same is true of many moral concepts, as moral

⁴⁹ The original reads: ‘Quelle vérité que ces montagnes borment, qui est mensonge au monde qui se tient au-delà?’ (1967, II.xii.241).

⁵⁰ I take the cocktail party example from Searle (2010, 33–34).

⁵¹ It notably fails to generalize to moral norms, for example; see Williams (2001b, 23).

realists rightly insist. Many concepts that are not universally used, much less universally used in an engaged way, nonetheless claim universal applicability and validity. And it is no surprise that we should have ways of thinking which, just because they are not relativized, at the level of their content, to our own dispositions, allow us think about people with different dispositions and express our concerns about them. Many normative concepts, including many moral, religious, or legal concepts, would be robbed of their *raison d'être* if they did not apply to those who did not already use them. It is the *point* of many such concepts that they apply to people who are not yet under their sway, and thereby enjoin alignment with the sensibility they themselves instil.

To present concepts claiming *universal* applicability and validity as claiming only *local* applicability and validity is thus to distort them by conflating two different aspects: the aspect under which concepts present themselves in disengaged use, and the aspect under which they present themselves in engaged use. The realization, from the disengaged perspective, that our reasons are indexed to our concepts is misguidedly inserted right into the *content* of our thoughts using the concepts in an engaged way. The thought: ‘This is a reason to φ ’ is reinterpreted as being, in effect, the thought: ‘For us, this is reason to φ ’. But that is a *different* thought. Compare the following two sentences:

(1) x 's being F counts as a reason to φ .

(2) For the set of engaged users of the concept F , x 's being F counts as a reason to φ .

The reasons adverted to by the concept F as understood in sentence (1) claim universal validity: even for those unacquainted with F , x 's being F is a reason to φ . The reasons adverted to by the concept F as understood in sentence (2), by contrast, only claim validity for a bounded set of reasoners, namely those who in fact use concept F in an engaged way.

Accordingly, it would result in a serious distortion of the contents we were trying to index to our own conceptual repertoire at the reflective level if we inserted the indexation into the contents of the concepts. This would be to let one's disengaged perspective on a concept distort the engaged perspective one takes up in using it—a distortion roughly equivalent to mistaking the normativity of morality for that of etiquette. Even once we realize that reasons which for us are ‘simply there’ are not necessarily ‘simply there’ for other people, we do not, unless we are changing the subject, suddenly have the thought: ‘For us, this is a reason’—we still have the thought: ‘This is a reason’. If, from the engaged

perspective, the reason is simply there, then the reason is properly understood from that perspective as applying *no matter what our concepts are*. That is *what it is* for a reason to be, for us, simply there; and so that is the thought whose dependence on our concepts we must grasp from the disengaged perspective without distorting it.⁵²

Much the same point can be put in terms of truth. Realizing, when using a concept in a disengaged way, that a truth articulable in terms of that concept is relative to our conceptual apparatus may tempt one to conclude that all we are really entitled to, when using the concept in an engaged way, is the thought: ‘It is true in our conceptual apparatus *A* that *x* is *F*’. But if what we are thinking in an engaged way is true in *A* and *A* is in fact the conceptual apparatus we are operating in, then what we are thinking is true *simpliciter*. Consider, by way of illustration, the analogy between conceptual norms and the rules of a game such as chess. If a player delivers a performance on the board which makes her the winner according to the rules of chess, it may well be that there is an alternative game, *quess*, in which the same performance would make her the loser; but when the time comes to hand someone the trophy, we have to decide whether it was chess or *quess* we were playing, and if it was chess, then that makes our player not merely the winner at chess and the loser at *quess*, but, simply, the winner.⁵³

At the same time, the fact that realist locutions correctly express our concepts as used from the engaged perspective does not entail that they offer us the best description of our concepts from the disengaged perspective. It would equally be a conflation of the two perspectives to maintain, in the face of the manifest variety of moral concepts across different cultures and epochs, and given the lack of a good explanation as to why only a tiny fraction of humanity ever arrived at the precise set of moral concepts in use now and around here, that those moral concepts are absolutely and definitively the right ones, because they are the only ones that pick out some metaphysically privileged properties that were crying out to be referred to all along. Within the disengaged perspective, there

⁵² See Williams (2006g, 195).

⁵³ Lewis (1983a, 173) makes this point about truth in a language. The illustration is adapted from Percival (1994, 191–192) and McPherson (2011, 232), though the chess analogy is of course frequently used by Wittgenstein, who remarks already in 1929 that ‘a pawn is the sum of rules for its moves (a square is a piece too), just as in the case of language the rules define the logic of a word’ (1975, 327–328). For further discussion of Wittgenstein’s use of the analogy, see Gustafsson (2020).

is room for the recognition that those concepts came to be ours through various historical contingencies; that we would now be responsive to different reasons had history been different; and that we would not necessarily be confused or deceived if it were so. It is merely that sentences making disengaged use of a concept to express perfectly correct philosophical-cum-ethnographic observations yield blatant falsehoods when interpreted as engaged uses of the concept, and vice versa.

Adopting the autoethnographic stance thus reveals a certain kind of dependence of reasons on concepts: the reasons we are responsive to are *hermeneutically dependent* on the concepts we use. The fact that $F(x)$ is a reason for $G(x)$ is hermeneutically dependent on concept F in that $F(x)$ would not be *intelligible* to us as a reason for $G(x)$ if we lacked concept F . But one must resist the unwarranted slide from this hermeneutic dependence to other kinds of dependence, such as the *ontological* dependence of *existence* on concepts, or the *logical* dependence of *reason relations* on concepts.

First, hermeneutic dependence does not imply *ontological* dependence, where some object or property P is ontologically dependent on concept F just in case P would not exist if we lacked concept F . As the example of the property of funniness showed, the fact that conceptual sensitivity to the presence of property P presupposes possession of the concept F does not entail that P cannot exist unless F exists. The slide from hermeneutic to ontological dependence leads to a crude idealism on which things only exist when we think of them.

Second, hermeneutic dependence does not imply *logical* dependence: the reason relations articulated by concept F are logically dependent on concept F just in case the reasons $F(x)$ adverts to are conditional, at the level of their content, on being an engaged user of concept F . But again, while it is true that $F(x)$ would not be intelligible to us as a reason for $G(x)$ if we lacked concept F , it does not follow that the kinds of reasons $F(x)$ adverts to are conditional on being an engaged user of concept F . That would only follow if F were a rather special kind of concept, namely one advertiring to reasons that were understood to apply only to engaged users of that concept. There are such concepts; but they typically involve enculturation into certain norms that are expressly understood to apply only to those who have been so enculturated; or else they involve what might be called the ‘knowledge *oblige*’ principle: the thought that, by acquiring certain concepts, one comes to see certain reasons whose applicability is conditional on being able to see

them. But it is clearly unwarranted to generalize from these highly specific structures to the conclusion that all reasons make their applicability conditional on engaged concept use in this way. The slide from hermeneutic to logical dependence leads to a jejune relativism on which reasons only ever apply to those who think in terms of them already.

What disengaged reflection on the contingency of our conceptual apparatus should prompt us to do is not to systematically recast the judgements formulated therein in terms that index them to us, in a misguided effort to be more truthful, but rather to ask whether we have good *reasons* to operate with this apparatus rather than another—*why* we play chess rather than quess. Perhaps things would go better for us in some respect if we played quess. Just as using alternative pieces guided by different rules would allow for different moves and combinations in chess, thereby changing the dynamics and character of the game, so the use of alternative concepts would allow for different inferential moves and combinations of thoughts.

1.7 Concepts for Conceptual Ethics

As Gilbert Ryle observed, it is easier to talk sense with concepts than to talk sense about them.⁵⁴ That was already true enough when Ryle wrote it in 1949. Since then, however, the difficulty has been compounded by the multiplication of theories of concepts in philosophy, psychology, and beyond, which has led to the term ‘concept’ being used in several tightly regimented but incompatible ways. It is therefore not unreasonable to insist that if we are going to reflect on concept appraisal, we need to specify what those ‘concepts’ are that are being appraised. At the same time, how we should conceptualize concepts depends on what we aspire to do with that conceptualization, which, in this case, is conceptual ethics. So how should we conceptualize concepts for the purposes of conceptual ethics?

I suggested that when appraising different ways in which people might think, we should consider not only what extensions people pick out with their concepts, but also what they take to be good indicators of something’s falling under a concept, and what they take to follow from it. This may already appear controversial, for while some philosophers are happy to recognize these further aspects of concept use as equally

⁵⁴ See Ryle (2009a, lx).

constitutive of a concept's *identity* conditions, many relegate them to a concept's *possession* conditions, constitutive not of the concept *F* itself, but merely of concept-users' *conceptualization* or *conception* of *F* (I use the terms interchangeably).⁵⁵ While it is widely agreed that concepts are the kinds of things we can possess, there is thus no consensus on the nature of the things possessed. How can conceptual ethics hope to get off the ground in the absence of a consensus on what concepts are?

The line I propose to take in response to this question is that the possibility of conceptual ethics, as I understand it here, does not depend on there being a consensus on the nature of concepts, because to ask which concepts we should *use* is to ask about the differences entailed by the *possession* of different concepts. To see the force of this line of argument, start from the observation that, while philosophers have proposed a variety of schemas for spelling out possession conditions,⁵⁶ there is broad agreement on the fact that concepts are the kinds of things that can be possessed, and that concept possession brings certain dispositions in its wake—not necessarily only dispositions that are automatically exercised in response to the presence of certain stimuli, but also dispositions one can intentionally refrain from exercising: abilities.⁵⁷

Notice that, if one thinks of concept possession as something relatively enduring that can actually manifest itself in some ethically relevant way in certain situations, one is forced to find room in one's account for abilities or dispositions to conceptualize things in the terms specified by a concept and be sensitive to the reasons it adverts to. As Hans-Johann Glock points out, 'identifying concept-possession with an ability or disposition of *some kind* is inevitable, and it is accepted, willy-nilly, even by proponents of RTM' (2006, 52). Jerry Fodor, for example, though a pre-eminent proponent of the Representational Theory of Mind (RTM), writes: '*having* a concept is: *being able* to mentally represent (hence to think about) whatever it's the concept of' (2003, 19, emphasis mine). Similarly, Robert Hanna, who holds that 'a concept is an essentially

⁵⁵ See, e.g., Cummins (2002, 88–89), Sainsbury and Tye (2012, 20), and Gauker (2011, 6).

⁵⁶ For different accounts of possession conditions, see Boghossian (2003), Williamson (2003), Fodor (2004), Glock (2006, 2009b, 2010), Eklund (2007), Scharp (2013, ch. 2), Hanna (2015, 62), and Glasgow (2020).

⁵⁷ See Kenny (1992, 66–85) and Glock (2000, 47). This is not to say that we can choose at will which concepts figure in a given judgement, however.

descriptive, more or less general, categorizing mental content with inherent linguistic and logical form' (2015, 60), nonetheless also concedes that if *X* is a concept, *X* is necessarily 'possessible', which he takes to entail that '*X* is deployable and usable, which is to say that *X* makes it really possible for cognitive subjects to recognize *X*-type things when they perceive them, ... to distinguish *X*-type things from other types of things', and 'to make analytically necessary and a priori logical inferences that pick out at least some of the intrinsic descriptive intensional elements of *X*' (2015, 60). On his view, concepts are emphatically *not* dispositions or abilities, but mental contents of intersubjectively shared types of mental representations that can be tokened by individual minds; and yet he still concedes that, on this view of the nature of concepts, concepts can be possessed, and concept possession confers certain discriminatory, classificatory, and inferential abilities. This is all that is required for conceptual ethics to get off the ground.

It is thus not particularly controversial to think of concept *possession* as a matter of having certain discriminatory, classificatory, and inferential dispositions or abilities. We attribute the concept *F* to someone on the grounds that they have the ability to locate and reidentify things that are *F* as opposed to non-*F*, to classify these things *as* being *F*, and to know what their being *F* implies.⁵⁸ We attribute the concept *red* to someone, for instance, on the grounds that they are able to do certain things, such as reidentifying red under different lighting conditions, distinguishing red from other colours, or inferring that if something is red, it is not green. Mastering a concept typically involves learning the techniques involved in exercising certain abilities. Such mastery comes in degrees: the expert concept-user might have more fine-grained ways of distinguishing *Fs* from non-*Fs* or a deeper understanding of the inferential ramifications of being *F* than the lay concept-user. Mastery of a concept also typically expresses itself in mastery of the use of a linguistic expression—Wittgenstein remarks that a concept is 'the technique of our use

⁵⁸ While classification and inference are standardly highlighted as key abilities involved in concept possession, the prior ability to reidentify or 'same-track' what one classifies and draws inferences about is helpfully foregrounded by Ruth Millikan, who writes that an 'animal's first job is to keep whatever part of its distal world it would learn about *in focus* ... to recognize Obama again from the front or the back or in a newspaper photograph or by his voice, to recognize tiger again given different views or sounds or kinds of spoor' (2017, 7).

of an expression: as it were, the railway network that we have built for it' (MS 163, 57r).⁵⁹ But even when their use patterns align, there are good reasons not to *identify* concepts with their verbal expression, nor to attribute concepts to others solely on the basis of their lexicon—our thinking techniques are one thing, our ways of expressing and conveying these techniques through language another, even though the two are interdependent in various ways.⁶⁰

But, for my purposes, concepts need not *themselves* be dispositions abilities; perhaps they are the norm-governed techniques that we employ in exercising these abilities;⁶¹ perhaps they are the abstract objects or mental representations involved in exercising these abilities. I can afford to remain agnostic on this, because the contribution to conceptual ethics I aim to make in this book is compatible with any of these accounts of the nature of concepts. An inquiry framed in terms of whether our concepts help us to live requires only the minimal assumption that the concepts which people in fact possess systematically *co-vary* with their discriminatory, classificatory, and inferential dispositions or abilities.

This is not to deny that the divide between those who regard the reasons that guide and flow from a concept's application as part of a concept's identity conditions and those who regard them as part of its possession conditions expresses substantial and interesting philosophical differences. Nor is it to deny that understanding the concept *F* as something completely external to human thinking, which captures the nature of *F* independently of whether it ever forms part of anybody's conceptualization, has a variety

⁵⁹ Similarly, Brandom attributes to Sellars the slogan that 'grasping a concept is mastering the use of a word' (2015b, 102).

⁶⁰ According to Bruno Snell's 'lexical principle', for example, something does not become 'an object of thought' until it is 'seen and known and designated by a word' (1953, 7). The methodological precept of using words as proxies for concepts—an approach sometimes adopted by historians in order to render their material manageable (see, e.g., Skinner 2009, 325)—can have the same effect as the lexical principle. Concepts are not just the shadows of words, and words make imperfect proxies for concepts. How else could we appreciate the challenges involved in putting a thought into words vividly brought out by Eli Alshanetsky's *Articulating a Thought* (2019)? For a battery of arguments as to why it is a mistake to identify language too closely with thought, and why the lexical meaning of a word should not be identified with the concept it expresses, see Sawyer (2020b) and Rieland (2022).

⁶¹ This is the account of concepts developed by Glock (2006, 2009a, b, 2010, 2020). Unlike rules, the norms governing thinking techniques need not be explicit, but can be implicit in our practices.

of virtues—including, notably, that it helpfully highlights how certain concepts (paradigmatically, concepts of natural kinds) function in a way that leaves concept-users beholden to the independent nature of what they pick out: they are concepts embodying an aspiration to discover how things are *anyway* that precisely acknowledges that some things are not exhausted by the ways we conceive of them—indeed, that our ways of conceiving of them might be almost limitlessly wrong.⁶²

But I do think that when extended to the thick normative concepts I focus on here, this hyper-externalist understanding of concepts creates awkward problems, leaving it mysterious where a culturally local, thick normative concept that is nevertheless also supposed to be entirely independent of individual and collective conceptualizations might draw its content from. In those cases, it seems more promising to resituate the concept–conception distinction within a less externalist spectrum, so that we can, from the disengaged perspective, make sense of those concepts in terms of human conceptions. Indeed, it is part of the methodologically humanistic approach to ethical, political, legal, cultural and aesthetic concepts pursued here that I prefer to understand concepts in terms of individual and collective ways of thinking rather than the other way round.

Even aside from that methodological preference in connection with thick normative concepts, however, there is a more basic reason why a focus on conceptualizations is appropriate for conceptual ethics: while the difference between those who explain concepts in terms of human conceptions and those who regard concepts as explainable completely independently of human conceptions does indeed mark a decisive fork in the road for certain questions in conceptual *engineering* (e.g. is it possible to change a concept?),⁶³ appraising concepts in terms of human conceptions must seem an

⁶² For a recent defence of an externalist metasemantic framework along these lines, see in particular Sawyer (2020c), which precisely contrasts mere subjective or communal conceptions with the externalist notion of a concept she develops in Sawyer (2018, 2020a, b), drawing on influential externalist treatments of natural kinds concepts by Burge (1979), Putnam (1973), and Kripke (1980).

⁶³ If concepts are taken to be Fregean senses, for example, the aspiration to engineer or change them may seem unintelligible; though the relevance of metasemantic questions about the nature of concepts to conceptual engineering of course depends on one's view of the latter; see, for example, Nado (2020) for a view of conceptual engineering which denies metasemantics a substantial role in it. Moreover, it would be

appropriate approach to conceptual *ethics* on either account. For conceptual ethics, as I understand it, just *is* the ethics of concept possession: the question of which concepts we should *use* is the question of which *conceptions* should be operative in our actual thinking processes—how we should *conceptualize* things.

We may of course still wonder what sustains or underlies our use of concepts, and this is where Fregeans peer up to the immutable denizens of an abstract realm above us and Fodorians peer down into the mental cogs inside us. But the question of which concepts we should use remains the same whatever the answer. Consider, by way of analogy, a movie running on a TV screen. We may wonder whether it is being streamed in from the world wide web or read off tiny crystals inside a DVD player. But the movie that plays on the screen is the same either way, and the question corresponding to the animating question of conceptual ethics is *which* movie should be playing. Analogously, we can make sense of the question of which concepts we have most reason to use whatever exactly is involved in using them. The Fregean can do conceptual ethics by asking which senses or modes of presentation should figure in the propositional contents we in fact articulate. The Fodorian can do conceptual ethics by asking which mental representations should govern our cognitive processes. What is more, each can grant that these choices will systematically co-vary with our dispositions and abilities. The animating question of conceptual ethics has force whichever of these metaphysical accounts of concepts one favours, because the question is not what is involved in possessing a concept, but which concepts we should possess.

It is thus appropriate for the approach I go on to develop here, as a contribution to conceptual ethics, to be methodologically agent-centred, focusing on identifying the conceptualizations—the concepts *as possessed, understood, and used by concept-users*—that best meet the needs of actual concept-users. This is not to *reduce* concepts to what we do with them, but rather to take what they do for us as a basis for appraising them. Needs-based appraisal will then be, in the first instance, an appraisal of conceptualizations, and any more external correlates these may have will be appraised only indirectly, via the conceptualizations through which they exert their influence on human affairs.

in keeping with the spirit of conceptual engineering to engineer whatever concept of a concept enables it to get off the ground—as Nado (2021a) herself does.

Accordingly, I use the term ‘concept’ to mean ‘conceptualization’ unless otherwise specified, and take it to refer not to abstract objects in an immutable realm nor to mental representations in the head, but to *ways of thinking and valuing*—patterns of thought most readily discernible in the weave of everyday life, at the level of what people pick up on, how they react, what they prove themselves capable of doing, and what trains of reasoning become manifest in their speech and behaviour.

In conceptualizing concepts as ways of thinking for the purposes of conceptual ethics, however, it is important to appreciate two respects in which they are neither purely psychological nor purely causal in character.

First, concepts are not purely psychological in that to use concepts is not primarily to think about concepts, but to think about the world and its objects. Moreover, the reason relations that concepts encode do not primarily obtain between psychological states, but between objective facts. It may be objected that it is the application of one concept that gives one reason to apply another concept, or that it is the judging that *p* that gives one reason to judge that *q*—and these all seem to be psychological operations. But we must be mindful here of what Wilfrid Sellars calls ‘the notorious “ing-ed” ambiguity’ between the acts of judging and the contents judged.⁶⁴ When I judge that I have reason to go to the dentist because I have a toothache, my reason is not *my judging* that I have a toothache, but the *fact* that I have a toothache, and what this fact gives me reason to do is not to sit there and judge things, but to go to the dentist. It is only in special cases that the act of judging is itself what gives me a reason—for example, when my judging that I hear the voice of Socrates speaking to me gives me a reason to see a psychiatrist.

In separating thinking from what is thought about, we need not deny that our acts of judging can also affect and reshape our concepts, however. We can accept that, as Robert Brandom puts it, ‘[o]ur judgments shape our concepts no less than our concepts shape our judgments’ (2015a, 17). This is the lesson of Quine’s critique of Carnap.⁶⁵ Carnap

⁶⁴ See Sellars (1997, §24). Thus, when Virgil writes of a crew competing in a boat-race: *possunt quia posse videntur*, ‘they can because they think they can’ (2007, V.231), it is the attitude of *believing* that they can win the race which makes true the proposition *believed*, namely that they can win the race. This is sometimes termed an ‘act-object’ or ‘state-content’ ambiguity (Alvarez 2010, 125).

⁶⁵ See, in particular, Quine (1960) and Carnap (1952). See also Brandom (2009, 83; 2014, 22–23; 2015a, 12–14; 2019b, 15) for especially lucid presentations of this Carnapian two-phase account and its replacement by a Quinean two-aspect account.

thinks of the determination and application of concepts as two separate and sequential activities: first we introduce and define our terms, exhaustively determining their proprieties of use in advance of their application; then, in a second phase, we apply those concepts to the world according to their stipulated proprieties of use, forming judgements that the world reveals to be either true or false. As radically free as we are to choose what concepts to use in the first phase, as rigidly bound by our concepts and by the world are we in the second phase. This two-phase model may fit artificial languages, Quine grants; but it fails to capture how, in natural languages, concept determination and application bleed into each other—indeed, *must* bleed into each other, since, outside formal logic, our use of a concept is not always preceded by a cool hour of careful stipulative definition. Sometimes, concept use is all there is, which means that determination and application then have to be thought of as two *aspects* of a single process of concept use. Far from fully settling the content of our concepts in advance of their application, it is often only through the process of deploying our concepts that we render them fully determinate. And far from treating our concepts as fixed and immune to their deployment in judgements, we sometimes revise our concepts through our use of them, by applying them in novel ways.

Second, concept use is not a purely causal affair in that it possesses a normative dimension that permits a distinction between correct and incorrect applications. Concept use is answerable to a standard of correctness, just as techniques can be mastered more or less well and are answerable to a standard of excellence. In that sense, ways of thinking are really thinking *techniques*.⁶⁶ The concepts I use are not simply to be equated with the sum of my actual dispositions to reason in certain ways, since talk of correctness becomes meaningless if whatever I am disposed to *treat* as correct *counts* as correct. This need not be taken to mean that I must be consciously guided by explicit norms as I apply a concept. But it does mean that I must be *subject to normative assessment* by third parties as I apply the concept. There minimally needs to be a communal practice of concept use to which the individual concept-user is accountable, and in relation to which a particular application of a concept can meaningfully count as correct or

⁶⁶ See Hacker (2013).

incorrect.⁶⁷

This accommodates the important point that, in grasping a concept, what one binds oneself to can outrun what one immediately grasps: what one immediately grasps, one's *individual* conceptualization, encompasses all those inferential moves one is disposed to make; but one's grasp of a concept may be only partial, and the concept whose proprieties of use one undertakes to respect by grasping it, however partially, encompasses all the moves one *would* make if one had a complete grasp of the concept.⁶⁸

Making sense of the nature of that normative dimension of concept use is precisely not my concern here, since my focus lies on authority of rather than in use. But the fact that concepts have such a normative dimension is nonetheless part of what motivates the authority question in the sense that concerns me: it is because a concept exercises authority *in use*, acting as a standard for correct application that users submit to by adopting the concept, that the question of the concept's authority *of use* is appropriate in the first place.

However, while this distinction between the authority of a concept and the normative dimension involved in applying a concept gives us the bare bones of an account of engaged concept use, it is not by itself sufficient to make sense of what engaged concept use is *sustained by*, or what in concept-users might *respond to* reasons for concept use. After all, it is not that the correct application of a concept is made *more* correct by the discovery of additional reasons to use the concept. To understand what reasons for concept use modulate, and what engaged concept use is sustained or undermined by, we need to flesh out our account of how concept-users relate to their concepts with another notion: the notion of confidence in a concept.

⁶⁷ Although concepts are usually shared, and there needs to be a social dimension to concept use for it to be liable to assessments of correctness, socially shared concepts can also become individualized, just as individualized concepts can become socially shared. Understanding the introduction and extinction of concepts requires a dynamic picture whereby the invention of a new concept whose use is initially not liable to normative assessments of correctness can, over time, become subject to assessments of correctness, and vice versa.

⁶⁸ For fuller articulations of a concept–conception distinction along these lines, see Brandom (1994, 9, 583, 632–36) and Wanderer (2008, 120–21).

CHAPTER TWO

Confidence, Reflection, and Knowledge

2.1 Confidence in Concepts

Engaged concept use is sustained, in the first instance, by confidence: the practical confidence with which we deploy concepts, without doubt, hesitation, or indecision, secure in the sense that these are the terms in which to cast our thoughts. This is confidence *in conceiving things in certain terms*—and, by extension, in the judgements articulable as a result.

Confidence is, in the first instance, the degree of our matter-of-factual attachment to our concepts. It manifests itself in our sense that our concepts are the right ones, in our willingness to rely on those concepts as a basis for our reflections and decisions, and in our receptiveness to the reasons adverted to by these concepts. Equally importantly, confidence manifests itself in how strongly a concept engages our emotions through the judgments it enables us to make. One's pre-reflective degree of confidence is a result of one's upbringing, socialisation, and personal experience, but it is also influenced by the opinions and experiences of others. One is more likely to have confidence in a concept if it is widely accepted and supported by people and institutions one trusts, and the work of cultivating a concept by keeping it in circulation, promulgating it, and reproducing the dispositions to think in those terms in subsequent generations, is necessarily a collective enterprise.

Fully confident users of a concept will not feel in the least uneasy about conceiving of things in those terms; they will show no indecisiveness, no tendency to second-guess

their own conceptual choices, no lingering doubt, even when the stakes are high and a lot turns on whether the concept they rely on is the right one. Their judgements will be full-throated, their emotional involvement wholehearted, their actions single-minded. This does not bar them from taking up a disengaging perspective on the concept; but to them, that disengaged use is likely to feel contrived and nugatory, as though someone focused on a glorious sculpture's shadow instead of the sculpture itself, since the force of the reasons revealed by conceptualizing things in those terms is not seriously in question. Confidence in a concept can thus be described as a special form of trust, whereby we put ourselves under the sway of a concept and relax into making engaged use of it, fully identifying with the reasons and commitments it engenders. To those whose confidence is shaken by reflection, however, it becomes a live question whether those reasons and commitments might not in fact be the real shadows—the shadows of our concepts.

Confidence does not by itself imply knowledge that these are the right concepts to use; nor does it imply normative entitlement to being confident.¹ But since concepts exert authority *in use* by defining norms of proper use that concept-users are answerable to, confidence can also be characterized in terms of one's relation to those norms. Fully confident concept-users understand these norms as binding on them. Less than fully confident concept-users, by contrast, struggle to make sense of these norms as binding on them: their acute consciousness of the fact that there are alternatives to these concepts, together with their perceived lack of reasons to regard these concepts as the right ones, undercuts the bindingness of those norms by casting doubt on whether *these* are the concepts they should let themselves be bound by. This loss in confidence will then be reflected in a corresponding weakening in the concepts' authority in use ('Since x is *blasphemous*, it follows that I have reason not to do it. But do I really?').

¹ My usage of the term 'confidence' differs in that regard from Miranda Fricker's (2000), who reserves the term for 'something we possess only if we are entitled to it' (97n18). Such terminological differences aside, however, I take myself to build on the substance of her discussion, which itself develops the notion of confidence that Williams often appeals to, but never thoroughly spells out; see Williams (1995*i*, 207–208; 1995*o*, 203; 2001*b*, 36; 2011, 189–190). See also Moore (2003), Hall (2014), Blackburn (2019), and Lukomska (2022) for valuable discussions. There are also parallels between loss of confidence and the Hegelian notion of *Entfremdung*, especially as interpreted by Brandom (2019*b*, 30, 472, 493–506), though the focus there is on the indiscriminate alienation from all norms induced by the 'modern insight into the role we play in instituting norms' (2019*b*, 30).

Confidence might thus also be characterized as one's hermeneutic relation to certain norms: it is a function of one's ability to make sense of a concept's authority in use as binding on oneself. To lose that ability in relation to a concept is to become unable to use the concept in an engaged way and regard the reasons it advert to as being, unequivocally, reasons for oneself. At the limit, when reduced to deploying the concept in a disengaged way, one can no longer identify with its correlative reasons and commitments. One can still use the concept in a disengaged way, understanding how the concept is correctly applied and what follows from its application. But one keeps the concept at arm's length and ceases to regard its reasons as one's own—not as a matter of choice, but because one is no longer *capable of making sense* of its reasons as reasons for oneself.

On this account, part of the task of those who put forward concepts for others to adopt—whether under the flag of conceptual analysis or conceptual engineering—is to foster confidence in the concepts they propose. And here it is important to recall that, for all the talk of giving reasons for reasons, this confidence can be achieved in a variety of ways, some of which follow a more clandestine and sinister script: people can be trained, lured, coaxed, cowed, or hoodwinked into being confident in certain concepts and kept ignorant of anything that threatens to unsettle that confidence.²

But in the context of liberal democratic societies shaped by Enlightenment ideals of rational autonomy and transparency, 'composers of definitions' are generally expected to eschew these more insidious and manipulative means of fostering confidence in concepts; they are expected to earn confidence through open, rational discourse, by articulating *reasons* to think in those terms. Even in liberal democratic societies, exceptions are made in certain domains, whose boundaries are themselves continually subject to ethical and political renegotiations: early education is a prime example, where the capacity to reflect critically about concepts is itself fostered only later, on the back of having been coaxed into accepting concepts uncritically at first.³ Yet, as a rule, and especially when it comes to reaffirming confidence that has been unsettled by reflection,

² I illustrate and expand on this point in Queloz and Bieber (2022).

³ On the question of where and on what basis these boundaries are to be drawn, see Queloz and Bieber (2022) as well as Queloz (2022a), where I draw on Williams's (1995j; 2002, 226; 2006j) ideas on education and the 'theory of persuasion'.

we tend to insist that confidence must be earned by giving reasons for reasons—and this insistence is itself an expression of our historical situation.

It is our confidence in concepts, then, that reasons for reasons modulate, and that engaged concept use is sustained or can be undermined by. Our confidence in concepts can be strengthened or weakened by reasons for concept use, because confidence, like trust, is in principle responsive to reasons. We strive to place our confidence in those concepts we have reasons to be confident in.

Let us therefore say that confidence is *vindicable* when a concept's authority of use can in fact be grounded in reasons for concept use, whether the concept-user knows it or not. And let us say that vindicable confidence becomes *vindicated* confidence when it is *shown* to the concept-user to be grounded in reasons for concept use. The difference between vindicable and vindicated confidence is thus epistemic: vindicated confidence is vindicable confidence that is *known* to be vindicable. Confidence is vindicable when placed in something authoritative, and vindicated when known to be so placed.

By questioning the authority of a concept, we ask whether it is worthy of confidence. Although the two are thus closely connected, confidence remains distinct from authority of use: confidence is a state directed at what one has confidence in, whereas authority of use is a normative property or status of that which *merits* confidence.

2.2 Metaconceptual Reflection

It is a truism that confidence can be undercut by reflection—the sportsperson who reflects on her motion can thereby momentarily lose the ability to perform it confidently. In the case of confidence in concepts, it is reflection *about* concepts that poses a threat to confidence: *metaconceptual reflection* on the concepts one uses and why one uses them.

Yet the threat to confidence in a given concept *F* does not have to take the form of a discovery made through engaged use of the concept—i.e. that ‘there is no such thing as *F*.’ This has been the dominant model of what can be wrong with a concept:⁴ as Sabina Lovibond describes the worries one might have about the concept *first-class mind*, for instance, they primarily express themselves in the form of growing doubts over where

⁴ See Richard (2008) and Mühlebach (2019, Forthcoming, manuscript) for critical overviews of the literature foregrounding this form of critique.

there *are* such things as first-class minds. But one's reasons for losing confidence in a concept need not always take the form of wondering whether the concept is *vacuous*, i.e. whether its extension is empty because the world we inhabit does not in fact contain any such thing. Nor need they take the form of wondering whether the concept suffers from *presuppositional failure*, so that one would have reason to resist judgements articulated in terms of the concept on the grounds that a crucial presupposition of their becoming candidates for truth or falsity remained unfulfilled.⁵ In both cases, concept-users discover, from within their engaged perspective on the concept, that the concept fails by its own lights: it makes its applicability dependent on conditions that are never in fact met, or makes presuppositions that never in fact hold.

But many concepts that do not fail by their own lights in either of those ways still have something wrong with them. The diagnoses of vacuity or presuppositional failure are rendered attractive by their terminal clarity: the verdict that a concept is vacuous, in particular, is satisfyingly unambivalent and final. We can simply put our foot down and insist that ‘There is no such thing as *F*’. But there can also be a kind of evasion involved in this. It may too comfortably cast as an epistemic error what is really an ethical failing demanding a more complex reaction. For even where the vacuity critique is directed at an appropriate target and successful, it leaves untouched the many alternative conceptions of the concept that do not suffer from the same vulnerability. Critiques of the concept of race on the grounds that modern genetics has revealed it to be vacuous, for instance, do nothing to undermine the many other conceptions of race that are simply too superficial to be vacuous, and call for a more ethical style of critique—a critique that cannot afford to ignore the reasons animating the use of these conceptions.⁶

At the most general level, however, the capacity of metaconceptual reflection to exert

⁵ See, e.g., Gibbard (1992) on why the concept *lewd* is objectionable; a similarly presuppositional account is articulated by Eklund (2017, 73), who writes that a normative concept is objectionable ‘iff, roughly, its use in some sense presupposes a false normative claim’. But see Richard (2008, 14–22) for an account of how these models fail to capture what is wrong with slurs, for example.

⁶ I elucidate the notion of conceptual ‘superficiality’ in Chapter 4 and put it to work in Chapter 8. For an example of a critique of the concept of race as vacuous, see Smith (2020, 53–62). For an account which proposes to replace the vacuous conception of race with three non-vacuous conceptions tailored to different sets of needs, see Hardimon (2017). Four different conceptions of race that do not fall prey to the vacuity critique are also articulated in Glasgow, Haslanger et al. (2019).

pressure on our confidence in concepts does not depend on there being anything especially wrong with them. We may still, if only episodically, lose confidence in them as a result of confrontations with alternative concepts. These can be perfectly innocuous concepts whose application conditions and presuppositions are straightforwardly satisfied when we deploy them in an engaged way. And yet we might still be brought to lose confidence in them when reflecting on them from a disengaged perspective.

For this to happen—for metaconceptual reflection to act as a drain on one's confidence in a concept—it is sufficient for the following three conditions to come together:

- (1) there have to be *real or notional alternatives* to the concept;
- (2) the reflective concept-user must be *aware* of these alternatives;
- (3) there must be a *perceived lack of independent reasons* to prefer one's own concepts over those alternatives.

The first condition is easily met. Many of the concepts we now use differ substantially from those at work in other societies, both past and present.⁷ That conceptual diversity already implies that people *could* think differently, because they *have* thought differently.

The mere fact of conceptual diversity is not by itself enough to affect confidence in concepts, however, since societies might in principle be entirely oblivious to this diversity. Of course, as a matter of actual history, few, if any, societies will have fit this description for any length of time. But in the *analytically basic* case that we may dub the *unreflective condition*, confidence goes unchallenged. Concept-users confidently deploy the concepts they acquired without engaging in any form of metaconceptual reflection. A society remains in the unreflective condition as long as nothing prompts its members to reflect critically on their concepts. To remain in such a condition for any length of time, a society would not only have to be maximally homogeneous in its concepts, in the sense of lacking conceptual diversity, but also completely insulated from other societies as well as totally lacking in representations of starkly different ways of thinking (in its historical

⁷ Some of the classic, if now somewhat dated, studies to that effect include Westermarck (1924), Benedict (1934), and Ladd (1957). For more recent work underscoring conceptual diversity and exploring its philosophical implications, see Shun and Wong (2004), Wong (2006), Ng (2024), and Cullity (forthcoming).

consciousness or cultural imaginary, for instance). The absence of doubt over the authority of concepts would then allow concept-users to be completely confident in their own conceptual outlook. They would not even so much as recognize it as peculiarly their own, because there would be nothing to contrast it with.

This is why, as the second condition registers, there needs to be something that generates *awareness* of conceptual diversity: there have to be *confrontations* with alternative ways of thinking. These can be real confrontations within a society that is conceptually heterogeneous, or between societies with different ways of life. But they can also be *notional* confrontations across time, when we come to represent to ourselves how differently people thought in the past, or even purely notional confrontations with imagined forms of life.

As a matter of fact, awareness of conceptual diversity is as old as history itself, at least in the sense of being evident already in the writings of Herodotus, the ancient Greek historian whom Cicero hailed as the ‘Father of History’. Herodotus recounts how Darius the Great, king of Persia, summoned the Greeks, who burned their fathers at death, and asked them for what price they would eat their fathers’ corpses. They retorted that there was no price for which they would do it. Darius then summoned the Callatiae, who were known to eat their fathers, and inquired for what price they would burn them instead. They implored the king not to speak of such horrors. Quoting the poet Pindar, Herodotus concludes that ‘custom is lord of all’ (1920, III, 38).

But even if the unreflective condition is just an ideal type, it offers a helpful starting point and a contrast foil to the condition that most societies are actually in. For the crucial fact about the unreflective condition is that, as long as a community is in that condition, confidence in concepts *does not require reasons*. Instead, *mere* confidence is enough to sustain the engaged use of concepts under which people can come to know what to think, what to do, and how to live. The concepts can do the work required of them—of providing structure, orientation, and meaning—on the basis of *nothing but* confidence.

Once concept-users are confronted with alternative concepts, however, this calls for metaconceptual reflection on the reasons they have to prefer their own concepts to those alternatives. The confrontation with alternative concepts prompts the realization that the reasons their concepts advert to—reasons that seem to be *simply there*—are really only simply there *for them*, while different reasons seem to be simply there for others. What

sort of reaction, if any, does that realization demand? In Amia Srinivasan's vivid dramatization of this perplexing question: 'What am I supposed to do with this other me, this shadow me ... who articulates the world in terms of concepts that are alien to my own? What if she is the right one, and I am the shadow?' (2019, 128).

What makes this question so disconcerting is that the very raising of it makes it harder to answer. Our confidence once shaken, more is needed to stabilize it again—metaconceptual reflection *raises the bar* for confidence. This is because the reflective confrontation with alternative concepts transforms the situation and imposes an additional epistemic burden: we no longer simply have to be confident in using our concepts; we now have to be confident in using *those* concepts *rather than* the alternatives. Confidence is sensitive to the presence of relevant alternatives.

In order to be justified in privileging our own concepts over those alternatives, we then need some understanding of how our concepts relate to those alternative concepts, and that understanding has to be such as to vindicate our privileging of our own concepts: it needs to offer *reasons* for us to regard the reasons adverted to by our concepts as being simply there. Mere confidence is no longer enough. We need to ground our confidence in reasons to think that our concepts *merit* confidence.

Of course, if we can immediately discern reasons why some concepts structure our lives while other societies get by perfectly well without them, we can still shrug off confrontations with alternative ways of thinking. When we realize that we use the concept of online privacy while people at the court of Charlemagne did not, for instance, this realization does little to shake our confidence in the concept, because the conceptual difference is readily explainable in terms of a story of technological development that simultaneously accounts for their lack of the concept and vindicates our continued use of the concept. In this case, our confidence is bolstered by an understanding of how our concepts relate to those alternative concepts that accounts for the divergence in such a way as to vindicate our sticking to our concepts. The same holds for some other thick normative concepts, such as *net neutrality* (which is given when internet service providers remain neutral between contents and do not privilege or block particular websites). But what about a concept such as *human right*, which came into wider currency in anything

like its present sense only in the 1970s?⁸

When we cannot immediately see the reasons why some concepts structure our lives while other societies do without them, the question whether this is the best way for us to structure our social world becomes acute. That is what the third condition—the perceived lack of independent reasons for concept use—registers. If that third condition is met as well, metaconceptual reflection truly puts pressure on our confidence in our concepts.

The same problem of confidence can arise also at a more fine-grained level, when we confront closely related alternative elaborations of one concept without apparent reason to prefer one over the other. We will then not wonder ‘Why use anything like this concept at all?’, but ‘Why use this particular conception in this context rather than another?’ As David Wong observes, ‘[t]he latter question can pose a question of confidence no less than the former’ (2006, 232).

What exactly is it that eats away at our confidence in concepts as a result of these confrontations? It is, one wants to say, the acute sense of the *contingency* of those concepts. But contingency is a slippery notion. To get a better grip on it, it helps to distinguish between *causal* and *rational* contingency.

A sense of *causal* contingency is induced when notional confrontations with our own past make us realize how easily our concepts might have been different, how close we came to going down the path towards a different conceptual order. Since contingency is the negation of necessity, our concepts then appear contingent in the sense of lacking causal necessity: the developments that brought them about were not bound to issue in those concepts, but might easily have taken a different course. This could be taken to mean that there is a possible world whose entire history up to the time at which the

⁸ See Moyn (2010), who argues that concepts foreshadowing the concept of a human right, such as the concept of a shared humanity, do not render the concept of a human right as understood from the 1970s onwards inevitable, but leave concrete legal questions so underdetermined that anything from early Stoicism through Christianity to the advent of human rights as a potential basis for infringing upon a country’s sovereignty is compatible with them. Far more sociohistorically local factors—such as the general dissatisfaction with the internal performance of new states that were granted sovereignty during the process of decolonization, or the state of American domestic politics in the 1970s—have to be drawn on to account for the existence of the concept in its present form (2010, 9–10, 15–17, 39, 69).

concept took hold is identical to ours, but where the concept failed to take hold;⁹ or it might be taken to mean only that certain salient proximate causes of the concept's taking hold are identical, and yet the outcome different.¹⁰

But it is clearly a version of this causal sense of contingency that the historian of political thought Quentin Skinner registers, for example, when he observes that grasping the 'sheer contingency' of 'the causal story' by which our concepts came to be ours creates a 'haunting sense of lost possibilities' that leaves historians 'almost inevitably Laodicean in their attachment to the values of the present time' (1994, 45). This half-hearted attitude towards our own concepts reflects an erosion of confidence.¹¹ When John Stuart Mill remarks in *On Liberty* (2003, 101) that what made a Churchman in London would have made a Confucian in Beijing, he is echoing a long tradition in medieval philosophy—as present in the works of Al-Dawwānī and Al-Ghazālī as in those of Peter Abelard and Ramon Llull¹²—of grappling with the sense of contingency precipitated by awareness of alternative faiths and the realization that one might easily have grown into different ways of thinking.

By contrast, a sense of *rational* contingency is induced by confrontations with real or notional alternatives when we see no argument for preferring the concepts we happen to have over alternatives—in particular, no *neutrally specifiable* argument, which is to say no reasons with the appropriate degree of independence from the concepts they are invoked

⁹ See Pollock (1984, 150–71) for a conception of causal contingency along these lines. It imposes a considerable epistemic burden on those who make claims of causal contingency, since they would need to show that the concept might not have taken hold even if everything that ever happened up until that moment had been identical.

¹⁰ See Zagzebski (1996, 119) for such a conception of causal contingency, which makes claims of causal contingency far easier to substantiate than Pollock's: 'To say that *S*'s doing *x* at *t* is causally contingent, we need not consider the entire past history of the world up to *t*, but only those events *C*1, *C*2 ... *C**n* that are part of the causal history of *x* at *t*. We ask if these events could have occurred though *x* itself does not ... when someone asks, "If the money supply increased, would the interest rates go down?", the question is not whether the rates would go down if everything that has happened in the past happened and the money supply increased. ... [T]o show that the fall of interest rates is causally contingent it would be sufficient if there is a world like this one in the causally relevant respects [in which the rates do not fall]' (Zagzebski 1996, 119).

¹¹ See also Srinivasan (2019) for a rich discussion of this theme.

¹² See Adamson (2022, 44–60).

to vindicate. Such independence is needed, since vindicating a way of thinking by reasons it itself instils, or by reasons articulated in terms of *coeval* concepts (i.e. concepts produced by the same forces that produced the way of thinking we are trying to vindicate), would be insufficiently neutral to count as more than self-congratulation.¹³ When no such reasons are forthcoming, our concepts appear contingent in the sense that they lack *rational* necessity: they fail to be necessitated by reasons. Rational contingency emerges not when we discover an unsettling connection to causes that might easily have gone the other way, but when a concept turns out to *lack* any connection to considerations vindicating it against alternatives.

Rational contingency is a deeper and more unsettling form of contingency than causal contingency. The realization that something is causally contingent loses its unsettling force if it is followed up by a demonstration that the thing in question is not rationally contingent. At most, we will count ourselves lucky that we made it to the place we anyway have good reason to prefer. But the realization that something is rationally contingent cannot similarly be allayed by showing that the thing in question is not causally contingent. That merely produces a sense that we were always going to end up with something we have no particular reason to prefer over alternatives.

Hence, what the authority question calls for, in response to faltering confidence under metaconceptual reflection, are reasons by which to dispel the sense of rational contingency. We need independent reasons to prefer our concepts over alternatives—and not just a historical narrative of causal inevitability—to ground the authority of a concept and vindicate our confidence in it.

Interestingly, *notional* as opposed to real confrontations with alternative ways of thinking are *more* rather than less apt to be a drain on confidence. This is because the less the difference in concepts makes a difference in practice, the more room we have to reflect on the significance of the fact that our concepts vary so radically. To the extent that a

¹³ An example of reasons for concept use that fail to leverage sufficiently neutral grounds because they appeal to coeval concepts would be the vindication of our confidence in the concept of liberalism in terms of the reasons for concept use provided by individuals' need for autonomy. This is perfectly coherent, but it remains too internal to distinctively liberal ways of thinking, since the concept of autonomy is itself a liberal concept, and individuals are only conceived of as having a need for autonomy given a liberal conception of the individual. For a related point directed at attempts to justify the liberal order, see Williams (2005j, 8).

confrontation with another way of thinking forces us to take a stand on some issue, we have to make engaged use of our own concepts to judge the issue, and this may include judging the other party and the way it thinks. But to the extent that we are not forced to resolve some practical question about how to behave towards the alternative we are confronting—which is typically, but neither necessarily nor only, the case in a notional confrontation—it becomes a *real practical possibility* to make disengaged use of our own concepts in thinking about the confrontation, thereby suspending engaged judgement. Then the gulf between the two ways of thinking and its implications for our aspiration to objectivity can become *the* fact about the situation. That is when conceptual diversity precipitates metaconceptual reflection on reasons for concept use.

The judgements suspended remain, for all that, as universalistic in their aspiration to applicability and validity as they were before, and it would be a mistake to think that reflective awareness of conceptual diversity must be registered by transforming judgements claiming universal applicability and validity into judgements claiming merely local validity and applicability—that would not be to acknowledge the parochiality of our universalistic concepts, but to exchange our universalistic concepts for a radically different set of concepts. We need to distinguish a concept's aspiration to universal applicability from the universal authority of the concept. As we shall see in Chapter 7, a universally applicable concept may only be locally authoritative. Either way, the content of our own concepts and judgements is not suddenly altered by confronting another way of thinking, whether notionally or in a way that calls on us to come to some practical decision. If we thought in universalistic terms before, we will continue to do so.

What varies with the nature of the confrontation, however, is the strength of the practical need to exercise our capacity to judge in these universalistic terms. It is that practical need that gives point to our taking a stand on the difference at issue. Of course, there is nothing to stop us from passing judgement on the other party even when there is no such need. But the judgement, unless it draws its point from being witnessed by some third party, will be at risk of being pointless.

Here, then, room opens up for a stance towards disagreement that is not primarily concerned to decide or resolve the disagreement one way or the other. Is it a *relativistic* stance? It is a stance that is certainly closely related, and possibly identical with, the only form of relativism that Bernard Williams, who is scathing about every other form of

relativism, allows for. He not only castigates ‘vulgar’ relativism as ‘possibly the most absurd view to have been advanced even in moral philosophy’ (2001b, 20), but also insists, more generally, that there is no room for relativism to tell us what normative judgements to make—it comes either too early, if the encounter with the relevant alternative outlook has yet to take place, or too late, if the encounter has taken place and there is already a need to decide what to do about it.

Yet there is nonetheless a ‘truth in relativism’, as Williams’s essay of that title acknowledges. He labels it ‘the relativism of distance’. The difficulty, however, is that Williams’s varying characterizations of what gives rise to this particular form of relativism—in terms of whether confrontations are real or notional, whether the outlooks one confronts are historical or contemporary, and whether they are real options for oneself—seem to pull in different directions, and make it hard to pin down what the decisive feature is that supposedly makes room for relativism.¹⁴ As a result, even sympathetic interpreters have tended to be somewhat puzzled by or even critical of Williams’s relativism of distance.¹⁵

To my mind, however, the fundamental distinction that provides the basis for the relativism of distance is a different one, though also one that is to be found in Williams’s text:

Some disagreements and divergences matter more than others. Above all, it matters whether the contrast of our outlook with another is one that makes a difference, whether a question has to be resolved about what life is going to be lived by one group or the other. (Williams 2011, 178)

The central contrast, which determines where and to what extent there is room for a kind of relativistic stance, is the contrast between disagreements we are under more practical pressure to resolve and disagreements we are under less practical pressure to resolve. That contrast may correlate strongly with whether a confrontation is real or notional and whether it involves contemporary or historical outlooks; but these correlated properties are not what rationally grounds the relativism of distance.¹⁶ It is perfectly conceivable for

¹⁴ See Williams (1981f, 142; 2003, 107–8; 2006i, 93; 2011, 180–85).

¹⁵ See Tasioulas (1998), Fricker (2010b, 2013), Lear (2011), Blackburn (2019), and Rini (2019).

¹⁶ Pace Ng (2019) and Gaitán and Viciana (2018).

there to be occasions on which real confrontations with contemporary outlooks put us under no pressure whatsoever to resolve a practical question. Conversely, it is equally conceivable for notional confrontations with past outlooks to put us under real pressure to resolve a practical question (when questions arise over what to do about the nasty views of ancestors, founders, or benefactors, for example). What fundamentally creates an opening for the relativism of distance is not the fact that a confrontation is notional or reaches across time, but the lack of practical pressure to resolve a practical question.

The idea that there is ‘no room’ for relativism will then hold true precisely to the extent that a confrontation with another way of thinking is a practical confrontation, which is to say one in which the situation requires us to take a stand on whatever issue brings two ways of thinking into confrontation with each other, and decide how we are going to relate and behave towards the other party. Conversely, to the extent that a confrontation is *not* a practical confrontation in this sense, there will be room for a kind of relativistic stance—room opened up by the absence of a practical need to deploy one’s normative concepts in an engaged way.

The distance that matters for the ‘relativism of distance’ is thus a matter of another outlook being, for whatever reason, far removed from our practical concerns in the sense that there is little practical pressure on us to take a stand on how we are going to relate to it. As Williams puts it, ‘moral outlooks will have a tendency to lose impetus if their expressions are not directed to people with whom one’s relations *need* to be regulated and defined’ (2006*i*, 93, emphasis mine). There is correspondingly little point to the engaged use of our concepts of moral appraisal in those confrontations, because the resulting appraisal will ‘lack the relation to our concerns which alone gives any point or substance to appraisal’ (Williams 1981*f*, 142).¹⁷ There will then be more to be gained from reflecting about what the difference in outlook can tell us about ourselves and our concepts.

But is this relativism? ‘One can call it a kind of relativism’, Williams writes, but

¹⁷ This is where Miranda Fricker, revising her earlier and more critical interpretation of Williams’s relativism, now locates the ‘essential motivating idea for the relativism of distance’ (2020a, 198), which I take to indicate a certain degree of convergence in our mildly revisionary readings of Williams: we both agree that what is really doing the work is whether moral appraisal is pointful in relation to our concerns in a given situation.

... it is very importantly different from what is standardly called relativism. Standard relativism says simply that if in culture A, X is favoured, and in culture B, Y is favoured, than X is right for A and Y is right for B; in particular, if ‘we’ think X right and ‘they’ think X wrong, then each party is right ‘for itself’. This differs from the relativism of distance because this tells people *what judgements to make*, whereas the relativism of distance tells them about certain judgements which they *need not* make. (2005d, 68, emphasis mine)

Standard relativism aims to arrive at substantive conclusions making engaged use of concepts. It endorses inferences *from* certain judgements making disengaged use of our own concepts and their alternatives *to* certain judgements making engaged use of those concepts. It thereby lays itself open to the charge of committing the mistake we identified in the previous chapter, of conflating engaged with disengaged use. The relativism of distance, by contrast, is more interested in what the judgements making disengaged use of concepts can tell us about ourselves and our concepts, and insists on the legitimacy of dwelling on this question when there is no need to make engaged use of our concepts.

Calling it ‘relativism’ is thus misleading. The fundamental point is rather that certain confrontations, namely those in which there is no need to decide and regulate relations between the two parties, are rightly seen as inviting metaconceptual reflection more than they invite engaged concept use resulting in moral appraisal. We might still be disposed to say that the others are wrong, but the more salient question will be whether they are *simply* wrong. Are they committing an error within a shared cognitive enterprise, or can we identify reasons why different reasons should seem to be simply there for them? Do we have reasons to use the concepts we do rather than theirs? And are the reasons for reasons we can discern on each side truly independent of the concepts whose use they purport to justify? This is the kind of disengaged metaconceptual reflection on our concepts that notional confrontations with alternative concepts are particularly apt to induce.

2.3 Knowledge under Concepts

Confrontations with alternative concepts can thus undercut our confidence in concepts. But in doing so, they may undercut more than just our confidence in concepts. They can also call into question the *knowledge* we thought we possessed under these concepts.

As epistemologists have pointed out, one might know that *p* in the absence of

alternatives, but cease to know that *p* once a competing hypothesis, *q*, is presented.¹⁸ From that point onwards, knowing that *p* additionally requires reasons to rule out *q*. The presence of relevant alternatives can raise the bar for knowledge.

Analogously, the knowledge that concept-users have under some concepts can be lost once they are confronted with alternative concepts. For these confrontations will induce metaconceptual reflection that is liable to unseat the concepts under which knowledge is possessed. It is thus not merely confidence that is sensitive to the presence of alternatives, but knowledge as well.

This is the line of thought that led Williams to his notorious conclusion that ‘reflection can destroy knowledge’ (2011, 163–164)—a claim that has, if anything, been met with even more bafflement and criticism from interpreters than his relativism of distance.¹⁹

Against these criticisms, I contend that the key to the claim that reflection can destroy knowledge is that it is a claim about how reflection can destroy the confidence required for engaged concept use. By losing the ability to use certain concepts in an engaged way, one loses access to the perspectival knowledge articulable under these concepts, but retains the ability to see, from a disengaged perspective, that what is thereby placed beyond one’s reach is knowledge—only no longer one’s *own* knowledge. This interpretation fits the summary that Williams later gave of his controversial argument:

Statements made by people using thick concepts that are not our own are not ... unintelligible to us. Nor can I see any reason for saying that they are, one and all, false. There seem to be perfectly good grounds for saying that some of them are what, in local terms, they are taken to be, namely true; and, since the people who use them satisfy other relevant conditions, we can say that those people have some knowledge under these concepts. But this is not knowledge that we share, since we do not share those concepts. One way, at least, in which such concepts go out of use is that people become more reflective about them; so some

¹⁸ See, e.g., Stine (1976), Goldman (1976), and Dretske (1981). A similar intuition has more recently animated contextualism in epistemology; see, e.g., Schaffer (2004), Blome-Tillmann (2009), and Ichikawa (2017).

¹⁹ For representative criticism, see Putnam (1981, 55), Moore (1991), Wright (1992, 38–39), Altham (1995), and Blackburn (2019). Moore (1997, ch. 8; 2003) offers a more sympathetic reading, as does Thomas (2006, 147–66), though they both still take issue with various aspects of Williams’s position. For a recent re-examination of Williams’s claim (which is also ultimately critical), see Rosen (2022).

knowledge, at least—that knowledge—can be lost under reflection. (1995i, 206)

Let us consider the argument in more detail. The thesis that reflection can destroy knowledge is meant to apply only on what Williams calls the ‘nonobjectivist’ model of ethical practices, on which we see ethical judgements as part of a way of life that a certain society has come to inhabit rather than as an attempt to get at the same set of ethical truths that other human societies, and perhaps even other creatures, are also trying to get at. On such a nonobjectivist model, judgements formed using thick ethical concepts (such as ‘*x* is *chaste*’) can yield ethical knowledge: these judgements are not merely capable of being believed and of being true; they are also world-guided in the right way to satisfy a number of different definitions of knowledge—they can be based on good reasons, have the right causal history, be produced by reliable cognitive processes, or track the fact that is their object across nearby possible worlds, for example.²⁰ Indeed, many thick concepts are so closely guided by easily observable empirical features of the world that one would expect users of these concepts to converge in their judgements, because the concepts leave them ‘nothing else to think’.

And yet that knowledge is vulnerable, according to Williams, to being destroyed by reflection—not, of course, in the sense that people continue to conceptualize things in the same way but come to view as *false* what they used to think *true*, since what they would have lost would then turn out never to have been knowledge in the first place; rather, the claim is that reflection can entrain the loss of knowledge by undermining people’s *confidence* in the concepts under which they had knowledge. ‘What I had in mind’, Williams clarifies in a later paper, ‘was the situation in which they no longer have the concept with which they used to express a certain class of beliefs. They lose a concept, and so cease to have a disposition that expresses itself in categorising the world in those terms’ (1995k, 238).²¹ To lose a concept is to lose a range of dispositions or abilities, and one of the abilities that can be lost that way is the *ability to know* what can be known under

²⁰ Writing in the early 1980s, Williams (2011, 158–164) relies on what was then one of the most sophisticated definitions of knowledge, namely Robert Nozick’s (1981, ch. 3) truth-tracking account. For an application of Williams’s model of ethical knowledge under thick concepts to Confucian ethical communities, see Ng (2024).

²¹ See Williams (1984, 223; 1995i, 208; 1995k, 238–239), and, for a discussion that aligns particularly well with the emphasis I give to Williams’s claim, see Williams (2010, 207–209).

that concept.

More precisely, the claim must be that reflection on a concept can bar people from continuing to use that concept in an *engaged* way, because, by shattering their confidence in the concept, it expels them from the group of engaged users of that concept. At the same time, people in that postlapsarian condition must retain the ability to use the concept in a *disengaged* way if they are to understand both that there really is knowledge to be had under that concept and that they no longer themselves possess it.

The engaged-disengaged distinction thus turns out to be crucial to making sense of the claim that reflection can destroy knowledge. If we accept that certain thick normative concepts can yield knowledge, confrontations with different casts of thought can, by destroying our confidence in those concepts, undermine our ability to use them in an engaged way, and thereby undermine our ability to know what can be known under those concepts. But as long as this leaves us able to use the concepts in a disengaged way, it still leaves us able to recognize that, for engaged users of those concepts, there is knowledge to be had under those concepts; only not for us, because we can no longer muster the confidence required to deploy these concepts in an engaged way.

Corresponding to this change is a shift from the autoethnographic to the ethnographic stance: initially, we have some concepts that we use in an engaged way, forming true judgements with them that are world-guided in the right way to count as knowledge. Thanks to our ability to take up the autoethnographic stance, we then engage in metaconceptual reflection on our reasons to use these concepts. If anything we find or fail to find in the course of this reflection drains our confidence in the concepts, we lose the ability to use them in an engaged way. We can still understand them and assess whether those who deploy them apply them correctly; but, though the concepts are part of our conceptual inheritance, they are no longer *our* concepts. We now stand to them as ethnographers stand to another culture's concepts that they understand, but do not live by. They may imaginatively identify with these concept-users—but to *imaginatively* identify with something is precisely *not* to identify with it. Consequently, the perspectival knowledge articulable under these concepts is no longer *our* knowledge.

Consider, by way of illustration, a young man who throws around a certain slur with the reckless abandon of a teenager. Assume that the judgements formed thereby are pieces of knowledge. As he matures, he comes to reflect on the concept, and discovers

forceful reasons not to use the concept. These need not be reasons to do with the emptiness of its extension. He may come to believe that this very way of thinking, this way of tying certain normative consequences to certain observable conditions, is objectionable on ethical or political grounds. This drains his confidence in the concept, leaving him unable to use it in an engaged way. He still understands the concept; after all, he used to live by it himself. But the propositions that can be known under this concept are no longer part of his own body knowledge—not because the propositions are now false or ill-justified (they are as true and justified as they ever were), but because, though he still understands what people who think that way are saying, he can no longer bring himself to think that way himself—not *in his own voice*, not in the engaged, confident way required for him to count the judgements he forms in those terms as part of *what he knows*. Having become estranged from the concept, he has also become estranged from the knowledge the concept puts within reach of its engaged users. In that sense, his estrangement from the concept bars him not just from a way of thinking, but from a *way of knowing*. Yet, just because he finds that way of knowing objectionable, he does not think of himself as having paid for his ethical gain with a significant epistemic loss. He no longer regards that kind of knowledge as worth having. If anything, he now regards it as worth not having.

2.4 Metaconceptual Knowledge

Metaconceptual reflection can thus make us lose knowledge by undermining confidence in the concepts in terms of which that knowledge is articulated. But can metaconceptual reflection also affirm knowledge by yielding metaconceptual knowledge that the concepts in question are the right ones? Can knowledge *under* concepts be reinforced by knowledge *about* concepts?

I will argue in Chapters 5–7 that it can. And here, I part ways with Williams, who holds that, at least where our thick normative concepts are concerned, we cannot have metaconceptual knowledge that they are the right ones. The best we can hope for is the ability to retain confidence in a concept even *after* metaconceptual reflection, on his view. But that confidence will not be vindicated by metaconceptual knowledge; it will be vindicated merely in the sense that no reasons to doubt or jettison the concept were identified under reflection. Thus, for Williams, for confidence in a thick normative

concept to be vindicated simply *is* for the concept to survive metaconceptual reflection, and it

... survives reflection just in the sense that we would not have encountered any considerations that led us to give it up, lose hold on it, or simply drift away from it, as modern societies in the past two centuries or less have, for instance, done one or more of those things in relation to the concept of *chastity*. While we shall have the knowledge that comes with the deployment of our surviving thick concepts, we shall still not have any knowledge to the effect that we have a definitively desirable set of such concepts. ... The thick concepts under which we can have some pieces of ethical knowledge are not themselves sustained by knowledge, but by confidence. (1995*i*, 207–8)

Where I propose to differ from Williams is in insisting that thick concepts *can* be sustained by knowledge, though not by knowledge to the effect that we have definitively desirable concepts: there can be metaconceptual knowledge that certain concepts are desirable *for* certain concept-users in certain circumstances. That does not make them *definitively* desirable, since both concept-users and their circumstances change. But what we shall have found, at the level of metaconceptual reflection on our reasons to use these concepts, will be knowledge, and that knowledge can offer rational support to the confidence that sustains those concepts.

The strong assumption that Williams accepts and I reject, then, is that metaconceptual reflection can offer *rational support* to one's confidence in thick normative concepts, and thereby *affirm* what knowledge one has under these concepts, *if and only if* it generates metaconceptual knowledge that these concepts are *absolutely and definitively* the best ones—and, by extension, that the social world they help to create is the best of all possible social worlds.

We might therefore say that Williams has a *non-relational* conception of the kind of metaconceptual knowledge required to affirm one's confidence in thick normative concepts. This requires knowledge that the concept *F* is best *simpliciter*:

The Non-Relational Conception of Metaconceptual Knowledge:

Metaconceptual reflection affirms concept-users' confidence in the concept *F* *iff* it yields metaconceptual knowledge that the concept *F* is part of the set of concepts that is absolutely and definitively best.

Since Williams does not think that such metaconceptual knowledge is forthcoming, he concludes that our engaged use of thick normative concepts cannot be sustained by metaconceptual knowledge, so that it must be sustained merely by confidence. Metaconceptual reflection can offer no rational support to confidence, on his view. It can help confidence only negatively, by not yielding any disparaging or incriminating revelations. The thick concepts that survive reflection, on this picture, are simply those that are left over once our conceptual repertoire has been cleared of archaic holdovers and ideological rot.

By contrast, I see no reason to think that our concepts need to be known to be absolutely and definitively the best ones. On the contrary: as we shall see in the next chapter, there are good reasons to think that aiming for the set of concepts that is absolutely and definitively best is a poor way of finding the concepts that actually serve us best, given our distinctive circumstances and problems. For metaconceptual reflection to offer positive rational support to our confidence in thick normative concepts, therefore, it is sufficient for it to show that these concepts are the best ones *for* those concept-users in that situation.

I thus propose to rely on a conception of metaconceptual knowledge that is *relational*, indexing the merits of concepts to the characteristics of concept-users and their circumstances, and requiring only knowledge that the concept *F* is best *for* those concept-users under those circumstances:

The Relational Conception of Metaconceptual Knowledge:

Metaconceptual reflection affirms concept-users' confidence in the concept *F* iff it yields metaconceptual knowledge that the concept *F* is part of the set of concepts that is best *for* those concept-users under those circumstances.

On this view, metaconceptual reflection can do more than leave one's confidence unscathed; it can offer positive rational support to one's confidence in certain concepts, and thereby affirm what knowledge one has under these concepts. This is the view I substantiate in the chapters to follow (where I also consider why such relationalism does not imply relativism).

In this chapter, however, my aim has been to clarify in what sense our object-level judgements are sustained, in the first instance, by confidence in our concepts. In the

unreflective condition, confidence can do this without rational support by reasons for concept use. But once the pandora's box of metaconceptual reflection has been opened, confidence comes to require second-order, metaconceptual judgements as to why these concepts *merit* our confidence. As soon as we enter the reflective condition, the prospect of sticking to one's concepts without being bothered by their contingency looks like an abdication of reason, a resignation to blind confidence, or worse, a ruthless embrace of bigotry.

Under conditions of modernity, where awareness of conceptual diversity is deeply ingrained in pluralistic and historically self-conscious societies, there is no going back to a less demanding, unreflective condition: both notional and real confrontations with alternative casts of thought are inevitable, and so are the epistemic burdens they place on us. At the same time, many of the formerly powerful means to make sense of the authority of concepts—in theocratic terms, for instance—have now lost the widespread allegiance they once commanded. We are thus both especially in need of, and particularly short on, means to make sense of the authority of our concepts. To explore this predicament and the complex response it requires is the task of the next chapter.

CHAPTER THREE

Escaping the Trilemma

In demanding reasons for reasons, we embark on the search for something in which to ground the authority of concepts. But what kinds of considerations should the authority of concepts be grounded in? In this chapter, I outline what I take to be the three salient answers on offer: *foundationalism*, *ironism*, and *holism*. I argue that these three answers are all unattractive, and that a compelling answer to the question needs to be able to escape this trilemma.

3.1 Generalized Foundationalism

The first horn of the trilemma can be broadly circumscribed as *foundationalism* about concepts. In the context of the present discussion, foundationalism about concepts can be understood as the project of identifying, on the basis of timeless rational foundations, a set of concepts that is absolutely and definitively best. For the foundationalist, the only truly authoritative concepts are those that can be validated as timelessly demanded by rational foundations that anyone has reason to recognize as such.

That definition deviates somewhat from the traditional foundationalism/coherentism distinction in theories of truth and justification, and it is too inclusive to be of much interest in its own right.¹ But its role here is that of a contrast agent, setting off the view

¹ For more orthodox articulations of the foundationalism/coherentism distinction, see Olsson (2017) and Hasan and Fumerton (2018).

to be advocated from the multifarious philosophical enterprises that have sought timeless rational foundations in one place or another: in an abstract realm of Forms; in a natural *telos*; in divine commands; in natural law; in the mind of God; in universal dictates of reason; or in basic structural features of the world. While the philosophers engaged in these enterprises were often not primarily bent on identifying authoritative concepts in particular, their views of rational authority nonetheless encourage certain ways of thinking about the authority of concepts rather than others. And what all variations of foundationalism about concepts have in common is that they view the fact that some concepts are *ours* as insignificant. The question about our concepts, as about any other concepts, past or possible, must be to what extent they approximate the set of concepts that is absolutely and definitively best. We might say that the characteristic aspiration of foundationalism about concepts is to transcend the concepts that are only contingently ours, and find the set of concepts that can be authenticated as absolutely and definitively best on the basis of timeless rational foundations of one sort or another.

The most enduringly influential form that foundationalism about concepts takes is the aspiration to find the set of concepts that ‘carve nature at its joints’. Plato introduced that rather grisly image merely by way of analogy, in the process of arguing for the reality of the Forms.² But the idea of a jointed nature became a guiding one for foundationalist projects in metaphysics and the philosophy of science. It still figures centrally in Eli Hirsch’s *Dividing Reality* (1993) and Theodore Sider’s *Writing the Book of the World* (2011), for example.³ The authoritative concepts, foundationalists of this stripe maintain, are the joint-carving ones, because nature possesses an antecedent structure that our concepts should reflect. These joint-carving concepts (e.g. certain concepts of physics) articulate a supervenience base for facts at higher levels of descriptions that are articulated in non-joint-carving terms (e.g. certain concepts of psychology). But these higher-level descriptions remain inferior approximations. As Sider insists, ‘there is a privileged way to “write the book of the world”’ (2011, 8).

Another metaphor sometimes used to articulate this ideal of fidelity to nature is that of ‘reference magnets’, where the idea is that certain parts of reality *attract* reference by

² See Plato (*Phaedrus*, 265e).

³ For a representative sample of perspectives on whether nature is ‘jointed’, see the essays in Campbell, O’Rourke, and Slater (2011).

our concepts: they are more eligible for reference than other parts in virtue of being metaphysically privileged in some way—they are ‘more natural’, or ‘more unified’, or distinguished in some respect that is supposed to be completely independent of human dispositions and concerns.⁴

Of course, timeless rational foundations may not bear on all kinds of concepts: the basic structure of reality, for example, may leave our aesthetic concepts underdetermined, so that it remains neutral between the aesthetic concepts of Baumgarten and those of the Bauhaus school. And perhaps the same is true of moral concepts or political concepts—or even of thick normative concepts in general.

But even today, foundationalism is by no means confined to the most basic and purely descriptive concepts of fundamental physics. In *Reality and Morality* (2020), for instance, Billy Dunaway extends the idea of reference magnetism to moral concepts, arguing that properties like ‘moral rightness’ or ‘obligation’ are objective, metaphysically privileged properties that our moral concepts are accountable to. Let us therefore say that foundationalism about concepts is *generalized* when it purports to apply not just to concepts in a specific domain, but to concepts across the board—including thick normative concepts—and across every dimension along which those concepts might earn their claim to authority.

Foundationalist pictures according to which the authoritative concepts rest on some immutable bedrock that remains undisturbed by historical and cultural shifts *can* make sense to people, as philosophy’s own history amply attests. But even where these pictures make sense, the foundationalist ambition to extricate our thought from its contingent circumstances runs up against the fact that a foundationalist picture’s plausibility itself remains conditioned by the natural, historical, and social setting in which it is presented: even when foundationalist theories have made sense to people, they have done so notably through the support of the sociohistorical setting in which they were put forward.⁵ This

⁴ The metaphor of ‘reference magnets’ is usually associated with David Lewis (1983b, 1984).

⁵ The point can be spelled out either in terms of the historical conditions that create a demand for a foundationalist picture—as Forrester (2019) and Smith (2021) do for the foundationalist elements in Rawls’s theory, for example—or, more fundamentally, in terms of the contingent dispositions and shared understandings and practices that ineluctably because constitutively condition the application of concepts (for all the controversies that Wittgenstein’s rule-following considerations have whirled up, some

indicates at least one respect in which foundationalism cannot achieve the total independence from contingent circumstances that it strives for, because it depends on mediation by contingent circumstances to make sense to people.

This limitation of the foundationalist ambition is important not because it necessarily vitiates the foundationalist enterprise, but because it reminds us to ask whether foundationalism can really make sense *to us*, under the circumstances that *we* find ourselves in. For once we press that question far enough, and confront it not just in connection with the concepts of elemental chemistry or particle physics, but also in connection with the thick ethical, political, legal, cultural, and aesthetic concepts that give texture and density to our social world, it becomes doubtful that foundationalism can make sense to us in a *generalized* form that would be authority-grounding (a) across the entire range of our concepts, and (b) across the entire range of dimensions along which concepts might earn their claim to authority.

This comes out when we consider the flipside of Sider's claim that concepts are authoritative when they carve at the joints, namely that all the concepts that do *not* carve at the joints are equally unauthoritative: there is nothing to set them apart, since the only thing that could set them apart is the property of being joint-carving, and they all lack *that*. To be sure, one could try to discriminate further here by introducing a suitably comparative correlate of the conceptual property of being joint-carving, such as the relation 'closer to being joint-carving than'.⁶ All non-joint-carving concepts would then be authoritative at least to the extent that they approximated the joint-carving concepts. Yet the underlying idea remains that the property of being joint-carving is the sole source of conceptual authority.

But clearly, this will not do as a *generalized* conception of conceptual authority. Once we turn our attention from the concepts of fundamental science to the thick normative

consensus has nonetheless crystallized around that broad implication—see the essays in Miller and Wright (2002) and Kusch (2006)). Something like this latter limitation is, I think, what Williams has in mind when he somewhat cryptically invokes the Wittgensteinian idea of the 'primacy of practice' to argue that 'foundationalism, even constructivist foundationalism, can never achieve what it wants', because it will 'seem to make sense ... only by virtue of the historical situation in which it is presented', and will 'always be subject to the condition that, to someone who is intelligently and informedly in that situation (and those are not empty conditions), it does or does not seem a sensible way to go on' (2005e, 25).

⁶ See Pérez Carballo (2020, 310–11).

concepts that distinctively structure different social worlds, the foundationalist account no longer rings true: can we really still believe that there is a timelessly and absolutely privileged way to write the book of the social world? These concepts seem to derive what conceptual authority they possess from some other basis. And even if it is granted that certain concepts of chemistry or fundamental physics are authoritative because they are joint-carving, there are many further dimensions along which concepts might earn their claim to authority.⁷ It is implausible to think that two concepts that failed to be joint-carving to exactly the same degree could not differ in some other way in their claim to being authoritative for us. And it is even less plausible to think that there could not be contexts in which concepts at higher levels of description—psychological, ethical, political, or legal concepts—would be just as authoritative, or *more* authoritative, than the concepts of chemistry or fundamental physics. In other words, even if some concepts are authoritative *because* they are joint-carving, it had better not follow that concepts that are *not* joint-carving are not authoritative. There has to be some other notion of conceptual authority that allows us to capture the sense in which joint-carving concepts are clearly inferior to higher-level concepts in understanding and navigating the complexities of human relationships, ethical appraisal, political debate, or legal argument.

If the foundationalist account no longer rings true for us when generalized from science to the thick normative concepts that structure different social worlds, this may be because, as the dust settles on post-modernism's campaigns to expose the contingency and parochiality of anything considered necessary or universal, we are, to an unprecedented degree, historically self-conscious not only in our thought, but also in our thought *about* thought. Awareness of the fact that the supposedly timeless rational foundations of the past have invariably been washed away along with the ideas that were built on them has become so deeply embedded in contemporary culture that we may find it especially hard to believe in timeless rational foundations. Not only are we no longer in the unreflective condition. We are also no longer in a reflective condition in which the

⁷ A point that also has a role to play in explaining why scientists develop the particular concepts of natural kinds they do—see, e.g., Laura Franklin-Hall on the ‘category influence hypothesis’—the claim that ‘the contours of scientific classifications are to some degree influenced by contingent features of scientists themselves’ (2015, 933).

anxiety induced by awareness of conceptual diversity could be alleviated by foundationalism. We know not just that concepts change, but that history is littered with unsuccessful attempts to rest the authority of concepts on ever different foundations. We know that, time and again through the ages, people have convinced themselves that they had succeeded in transcending their contingent circumstances, only to discover later that the veneer of timelessness had itself been but a product of its time. Awareness of this fact is now too pervasive for us not to expect further attempts to identify timeless foundations to suffer a similar fate, and end up saying more about us than about the semipiternal scaffolding of thought.

3.2 Indiscriminate Ironism

Where does disillusionment with generalized foundationalism leave us? It might seem that the only reasonable response, once it is granted that we cannot vindicate any one set of thick normative concepts against alternatives on the basis of timeless rational foundations, is *ironism*: the sort of ironic detachment and disengagement from our concepts that Richard Rorty identifies in *Contingency, Irony, and Solidarity* as the only appropriate attitude for those who abandon the hope that their concepts could be rested on some bedrock ‘beyond the reach of time and chance’ (1989, xv).⁸

Ironism is the second horn of the trilemma. Without rational foundations from which to derive a set of concepts carrying the stamp of authority, there seems to be no independent reason to prefer one set of concepts over another. Irony, as Gideon Rosen puts it, is a second-order attitude that one comes to adopt towards one’s first-order judgements once it becomes clear that others may resist using one’s concepts ‘without making any neutrally identifiable mistake’ (2022, 152). It is a second-order attitude, we might add, that one adopts in lieu of second-order reasons to consider one’s concepts to be absolutely and definitively best.

On Rorty’s characterization, ironists have three characteristics: (i) they have radical doubts about their own concepts, because they are impressed by the alternative concepts

⁸ That introductory characterization of the ironist is couched in terms of beliefs rather than concepts, but the more careful characterization in a later chapter is couched in terms of choosing vocabularies (1989, 73), for reasons discussed in Santelli (2020).

they encountered in the lives of other people or in books, and by how final these alternative concepts were taken to be; (ii) they realize that arguments cast in terms of their own concepts can neither underwrite nor dissolve these doubts; (iii) insofar as they reflect about their situation, they do not think that their concepts are closer to reality than those of others, and view the choice between concepts as neither decidable within a set of neutral and absolute concepts nor decidable by fighting one's way past appearances to the real.⁹

Forced to continue to live by some concepts in practice, ironists may hold on to their entrenched concepts. But they do so out of ungrounded solidarity with 'the tribe'. In private, their reflective awareness of the rational contingency of their concepts prevents ironists from fully identifying with those concepts. As we can now put it: they lose confidence in their concepts and are restricted to using them only in a disengaged way.

Although the figure of the ironist has attracted much criticism,¹⁰ it registers something important which it would be complacent to dismiss, namely that reflective awareness of contingency does not leave everything where it was. Once 'we become conscious of ethical variation and of the kinds of explanation it may receive', Williams writes, it is simply 'incredible that this consciousness should just leave everything where it was and not affect our ethical thought itself' (2011, 177). Reacting to this consciousness by trying to derive from it 'a nonrelativistic morality of universal toleration' is 'seriously confused' (2011, 177), Williams grants, but even a confused reaction is a reaction to something, and it at least has the merit of acknowledging that some reaction is called for.

Certainly, closing our eyes to the consciousness of contingency and going on as before should not be mistaken for a return to the undemanding confidence of a society as yet untouched by reflective awareness of contingency. For once ignorance of contingency has been dispelled, it cannot be reestablished. As Thomas Paine insisted in a phrase that now carries Rawlsian resonances: 'when once the veil begins to rend, it admits not of repair ... though man may be *kept* ignorant, he cannot be *made* ignorant' (1998, 167). For us, pretending that *mere* confidence in one's concepts could be enough would amount to complacency bordering on chauvinism.

The attitude of the ironist thus has the virtue of being open-eyed and honest: it

⁹ See Rorty (1989, 73).

¹⁰ See, e.g., Blackburn (1998, 288–294), Fricker (2000, 2013) and Rosen (2022).

recognizes that, in the wake of metaconceptual reflection, confidence needs to be grounded in some basis for discriminating between concepts that merit confidence and concepts that do not.

Yet even ironism remains, in one crucial respect, foundationalist—notwithstanding the fact that it precisely denies that foundations of the sought-after kind are available. Ironism remains *counterfactually* foundationalist, because while it acknowledges that foundations are not in fact available, it still holds on to the foundationalist idea that if the authority of any concepts *could* be vindicated, this vindication would *have* to take the form of showing that the concepts were derivable from timeless rational foundations.¹¹ The ironist thus continues to endorse the conditional: ‘If a concept is authoritative, then it is authoritative by virtue of being derivable from timeless rational foundations’. It is the pattern of reasoning expressed in this conditional which licenses the ironist’s signature move from the observation that there are no foundations to the conclusion that no concept is really authoritative. But to reason that if no foundations are available, any concept is as good as any other is to remain attached to a foundationalist conception of conceptual authority—much as atheists retain a theistic conception of morality if they endorse Ivan’s inference, in *The Brothers Karamazov*, that if God does not exist, everything is permitted.

A similarly counterfactual form of foundationalism animates the view of a precursor to Rorty’s ironist, namely the early Nietzsche, who, in some passages, dismisses all human concepts as equally distorting and falsifying, because they invariably fail to correspond to things as they are ‘in themselves’.¹² In these passages, Nietzsche accepts that no

¹¹ I take the widely applicable and useful idea of *counterfactual* adherence to a position from Williams’s discussion of what he calls the ‘counterfactual scientism’ (2006g, 187) of the ironist. See also Queloz and Cueni (2019) for a different application of the idea to Nietzsche’s critique of asceticism.

¹² In addition to this Neo-Kantian line of objection to our concepts he presses notably in ‘On Truth and Lie’, Nietzsche also suggests that all our concepts systematically obfuscate differences: we use the same concept for ‘countless more or less similar cases which, strictly speaking, are never equal’ (TL 256). But this line has its own difficulties. To apply the concept *leaf* to two different things is not to claim that they are identical with each other, but only that they are similar to each other in certain respects. When we do want to make claims of numerical identity, we achieve this not by applying *the same concept* to the leaf, but by applying *the concept of the same leaf*. Moreover, the fact that most concepts have ‘countless more or less similar cases’ as their extension need not be taken to compromise our capacity to describe the individual

foundations from which to assess the authority of concepts are in fact available to us; but he holds on to the foundationalist thought that if it *were* possible to vindicate the authority of any concepts, that vindication would have to take the form of a derivation from such foundations.

Pace Williams (2002, 17), the early Nietzsche need not be read as holding that we *can* in fact look round the edge of all our concepts at the True World we are applying them to, grasp its nature without drawing on any concepts, and use that as a rational basis on which to appraise our concepts. But even in suggesting merely that the set of concepts corresponding to the True World is what, *per impossibile*, we would really like to have, and that our concepts are the less authoritative for falling short of that ideal, the early Nietzsche still clings, if only counterfactually, to a foundationalist conception of conceptual authority. Foundationalism then gives way to an indiscriminate disengagement from all concepts. As Nietzsche himself later diagnosed the bind he was in, being ‘freed from the tyranny of “eternal” concepts’ puts one at risk of ‘plunging into the abyss of a sceptical indiscriminateness’ (85:35[6]).

Foundationalism and ironism thus share the same conception of conceptual authority, and the same conception of reasons for concept use. They both hold that the only reasons that can properly *count* as second-order reasons for our first-order reasons are those that are reasons *for anyone*—for any concept-user, whoever they are, and whatever cultural context they find themselves in—and mark out one set of concepts as absolutely and definitively best.

We might say that these first two horns of the trilemma share a normative expectation concerning what second-order reasons *ought* to be available. What separates them are their empirical expectations concerning what second-order reasons are *in fact* likely to be available. The ironist’s normative expectation is disappointed by her empirical expectation. This is what creates the sense that something crucial is lacking, entraining indiscriminate disengagement from her concepts.

But foundationalists and ironists agree in what they regard as a good answer to the authority question: if we are to vindicate our confidence in any concepts, conceptual

case in full detail. Concepts can be combined to form indefinitely fine-grained descriptions, taking us from ‘A leaf again’ to ‘At location L and time t , there is a leaf with properties $P_1–P_n$.’ In fact, Nietzsche’s argument presupposes the ability to conceptualise the differences that concepts allegedly obfuscate.

authority must be grounded in reasons that are reasons for anyone because they derive from an absolute standard.

3.3 Reasons for Us: Non-Foundationalism

We need not choose between foundationalism and ironism. Once we see their shared presupposition, it becomes clear that we can break new ground by questioning it. Once we abandon not only the search for timeless rational foundations, but also the very hankering after such foundations, we go one step further in the recognition of contingency than the ironist: we acknowledge that contingency extends to the very *standards* that concepts are beholden to. It is not just the concepts that are contingently ours, but also the demands they must meet. We should not seek the concepts that are absolutely best, in that *anyone* has reason to use them. We should seek the concepts that *we* have reason to use, given the demands on our conceptual apparatus in the sociohistorical setting that is peculiarly ours. Parting ways with both foundationalists and counterfactual foundationalists, we then seek some *non-foundationalist* answer to the authority question.

The mark of non-foundationalist answers to the authority question is that they are given in terms of reasons *for us* that are not necessarily reasons for anyone. The key difference lies in what one is prepared to count as a normative resource sufficient to vindicate the authority of a concept. For foundationalists and ironists alike, the only truly authoritative concepts are those that can be validated as timelessly demanded by rational foundations that anyone has reason to recognize as such. For non-foundationalists, conceptual authority can be vindicated by standards that are less than universal. They see no reason to think that the best concepts for citizens of twenty-first century constitutional democracies, facing such entirely unprecedented challenges as the climate crisis or the risks posed by new digital capabilities, would be the same set of concepts as for medieval monks or Bronze Age chieftains.¹³ If the problems we face today are unprecedented, it would be surprising if the concepts that proved most helpful in tackling them were unresponsive to that fact. It is not just concept-users that are sociohistorically

¹³ On the various challenges posed by new digital technologies, see notably Zuboff (2015, 2019), Susskind (2018), Nemitz (2018), Macnish and Galliott (2020), Vélez (2020), and Smith (2021).

situated, but also the normative standards that their ways of thinking must meet.

To move beyond foundationalism and ironism, then, we need to free ourselves of the very urge to move towards a set of concepts that is entirely free of local perspective—what Amia Srinivasan calls ‘the Archimedean urge’ (2015). Our concepts are not all equally perspectival. Some concepts reflect the parochial sensibility and sensory peculiarities of their users more than others, and are more closely tied to a distinctive cultural and physiological perspective. When Blaise Pascal described the universe as immense, silent, and terrifying, for example, he was speaking from an increasingly local perspective.¹⁴ And once we realize that our concepts and descriptions are to varying degrees tinged with subjectivity, we may form the aspiration to wring ourselves out of our concepts and arrive at a description of the world that is fully objective.

Admittedly, this Archimedean urge to filter out the respects in which our conceptualizations of the world reflect our perspective does make sense in certain contexts—in physics, notably, where it counts as a scientific advance to move from using ordinary colour concepts like *red* and *green*, which are tied to a particular physiological perspective, to concepts like *wavelength* and *frequency*, which are accessible in principle to a much broader constituency. In the context of this kind of inquiry, where the inquirer precisely aims to shed the parochiality of his or her cultural and physiological perspective and approximate the Cartesian ideal of the ‘Pure Enquirer’, as Williams calls it, we really do have reason to strive towards concepts that are minimally perspectival—concepts enabling us, at the limit, to articulate a ‘scientific representation of the material world’ that is ‘absolute’ or non-perspectival in that it ‘does not have among its concepts any which reflect merely a local interest, taste or sensory peculiarity’ (2005b, 229).¹⁵

¹⁴ I take the example from Williams (2014c, 323). For a systematic elaboration of this notion of a perspective, see Moore (1997). As Moore emphasizes, the absolute/perspectival distinction applies exclusively to our representational devices, including notably our concepts, but not to what is represented thereby (1997, ch. 3). Some representational devices are inherently perspectival, and representations cast in those perspectival terms can be directly endorsed only if one shares the relevant perspective (though they can be indirectly endorsed by endorsing a representation that entails them); but all representations remain representations of one and the same world. See Moore (1997, 16, 35, 49).

¹⁵ This is one of the senses Williams gives to ‘Pure Enquiry’ in his discussion of Descartes (2005b, 49). It is a project Williams staunchly resists in the ethical sphere, where he is critical of the ambition to take up a Sidgwickian ‘point of view of the universe’ (1995h, 170; 2003; 2011, ch. 2).

The notion of a non-perspectival concept may strike one as incoherent at first: any concept that someone actually uses will have to be used from some perspective or other, and there will inevitably be some respect in which the use of the concept will end up betraying the perspective from which it is being used. In theoretical physics, for example, even the most abstract concepts are likely still to betray the perspective of their users in various respects—through the notational conventions they are expressed in, for one thing, or through the shared practice of interpreting the notation as expressing those concepts, or through the shared dispositions to apply the concepts to new cases in the same way.¹⁶ But, as Moore argues, a representation that *betrays* a perspective need not, for all that, be a representation *from* a perspective (1997, 89); even if concept use always bears the imprint of concept-users, it does not follow that the concepts used are always perspectival concepts, i.e. concepts whose conceptual *content* is perspectival. Concepts might betray a perspective without being perspectival concepts—just as a tenseless sentence, indexing an event to a specific date without making use of any temporally perspectival concepts such as *yesterday* or *one year ago*, might nonetheless betray the temporal and linguistic perspective of its utterer.¹⁷

We can therefore make sense of the aspiration to arrive at a description of the world couched in terms of non-perspectival concepts by acknowledging, first, that its intelligibility requires no more than the possibility of *approximating* the non-perspectival by replacing *more* perspectival with *less* perspectival concepts; and second, that even genuinely *non*-perspectival concepts are not in fact all that hard to come by, since concepts can be non-perspectival concepts even when their use still betrays a perspective.

And yet, despite the long-standing ‘ideological alignment’ (Moore 2020, 129) between the scientific, the absolute, and the authoritative, it would be a mistake to conclude from this that concepts which are closely tied to a local perspective are inherently and generally inferior to concepts which are less closely tied to a local perspective—that, as we might put it, the authority of concepts is inversely proportional to their perspectivalness. That would be to overgeneralize a model of conceptual

¹⁶ See Putnam (1992, 94–99; 2001; 2002, 40–45).

¹⁷ See Moore (2020, Forthcoming) for a defence of the possibility of absolute or non-perspectival concepts against recent criticism, including by Rödl (2018, ch. 5), and Moore (2019a, c) for an application of these ideas to tensed representations.

authority that has its place within the scientific enterprise, but that would be misleading elsewhere. The less aligned with scientific concerns our concerns in a different context are, the less clear it is that we should try to wring ourselves out of our concepts.

In the context of ethics or politics, for example, it is not even an intelligible aspiration to try to arrive at a non-perspectival description of the world that even alien thought might converge on. Here, as George Eliot puts it in *Daniel Deronda*, ‘that bird’s eye reasonableness which soars to avoid preference … loses all sense of quality’ (1999, 814). Here we are, on the contrary, trying to grasp how a situation relates to *our* concerns, attachments, and loyalties. Bringing those to bear on our judgements is not, in the first instance, to distort our view of the situation, but rather what gives evaluative contour and colour to the situation in the first place. Just as one could not even begin to choose how to live from the utterly unconcerned, detached, and indifferent point of view of the universe, one could not even begin to choose which thick normative concepts to live by. In the first instance, the choice between the concepts that give us reasons for action and structure our ethical, social, and political lives is not biased by our local perspective, but constructed and concretized by it, because that is the perspective that fleshes out *for whom* the concepts in question are to be authoritative. It is only from within an evaluative perspective that one can intelligibly strive to correct for bias and attain some degree of impartiality. Impartiality is not the absence of an evaluative perspective, but a value bearing on how to evaluate from such a perspective.

The choice of ethical, political, or indeed social or aesthetic concepts has to reflect the constitutively perspectival nature of these forms of thought. In these contexts, bringing our concerns, attachments, and loyalties to bear on concept choice is not, in the first instance, a distortion, but rather what makes it possible for the choice to be based in reasons instead of being arbitrary. This is why Williams cautions against the ‘scientific illusion’ that it is our task as rational concept-users to ‘search for, or at least move as best we can towards, a system of political and ethical ideas which would be the best from an absolute point of view, a point of view that was free of contingent historical perspective’ (2006g, 193–194). The thought that concept choice could be biased by local concerns is important, but it needs to be brought in later, once there *is* a choice to bias. Consequently, the Archimedean urge has no place here. Here, we should not first determine which concepts are absolutely and definitively best, and then infer from this that they should be

used from a perspective that happens to be ours. Here, the choice of concepts is not just incidentally, but essentially ours.

For non-foundationalist approaches to the authority question, then, the reasons we might have to use certain concepts can properly be reasons *for us* without being reasons that anyone would recognize as such. This need not mean that the reasons *in terms of which* we have reason to think are themselves reasons for us as opposed to anyone. The force of second-order reasons can be indexed to a particular perspective or constituency without thereby similarly indexing the force of the first-order reasons they are reasons for: perhaps, as numerous historians of the concept of authenticity have suggested, ‘we moderns’, living in far larger social structures that do less to predetermine what professional roles and group affiliations the individual will identify with, have particular reason to think in terms of *authenticity*;¹⁸ but the concept of authenticity we have reason to use nonetheless involves the idea that *anyone* has reason to prefer an authentic over an inauthentic life, not just that we moderns do.

The possibility of indexation to a perspective or constituency at the level of second-order reasons does mean, however, that, for each concept, we need to take seriously the question of who ‘we’ is: who has reason to use it and who does not. In most cases, the ‘we’ used in the present book is not the all-encompassing ‘we’ quantifying over all rational beings or all of humanity. Nor is it the presumptuous ‘we’ that imputes certain characteristics to an antecedently designated set of people—to which the reader is blankly presumed to belong—and tells them how to think. It is, rather, the ‘we’ of *invitation*, which invites readers to consider to what extent they recognize themselves in a certain description, and what this would imply.¹⁹

¹⁸ See Guignon (2004, 17–19), Lindholm (2013, 365), Taylor (1989, 26), and Trilling (1972, 15–16); though, as historians of earlier societies tend to emphasize—see, e.g., O’Doherty and Schmieder (2015)—the dynamism and mobility of medieval societies, for instance, should not be underestimated. See Leuenberger (2021) for a nuanced philosophical assessment of this historical material that highlights the practical reasons to live by the concept authenticity under conditions of modernity.

¹⁹ I am influenced here by an endnote to *Shame and Necessity* in which Williams justifies his own copious use of ‘we’ with the remark that it ‘operates through invitation’: ‘It is not a matter of “I” telling “you” what I and others think, but of my asking you to consider to what extent you and I think some things and perhaps need to think others’ (1993, 171n7). This conception of the pragmatics of philosophical texts is a natural

This construal of ‘we’ as functioning through invitation rather than through an antecedently fixed designation offers a plausible interpretative model for much philosophical writing, especially in ethics: many philosophical texts do not simply take for granted the sweeping generalizations about people that they use as premises; nor is their capacity to persuade passively dependent on a set of immutable and fully transparent concerns that the reader happens to bring to the text. Our sense of what our own concerns are is incomplete and malleable. Much philosophical writing, responding to this fact, aims to be persuasive in a more active sense, by striving to shape the reader’s sense of her own concerns and their relative importance. Insofar as these concerns do then come to be invoked as premises in the text’s arguments, they are concerns that might be *made* into important concerns for the reader through her engagement with that very text: persuasive philosophical writing can awaken, revivify, or strengthen concerns in the reader, and thereby contribute to the satisfaction of its own presuppositions.

Similarly, in speaking of ‘reasons for us’, I mean those of us who share certain reason-giving characteristics, where the idea is not to ascribe these characteristics to an independently identified set of people, but to invite the reader to consider to what extent they identify with those characteristics, and to what extent that might give them reasons to prefer certain concepts over others—reasons that might be reasons *for us* in highly localized sense.

3.4 Undiscriminating Holism

Once we opt for a form of non-foundationalism, we thus look to local standards to ground conceptual authority. But the way in which this has typically been done in twentieth-century analytic philosophy is through some form of *holism* about one’s web of concepts. This is the third horn of the trilemma.

The influence of holistic non-foundationalism in twentieth-century analytic philosophy is due notably to Rudolf Carnap’s voluntarism about framework choice and

consequence of Williams’s view that the reader’s thought ‘cannot simply be dominated’, because, ‘as it used to say on the packets of cake mix’, the reader will and must ‘add his own egg’ (1986, 203) in making something of the text. In this regard, Williams was self-consciously echoing Nietzsche and especially Collingwood (Williams 2006b, 343–4). For further discussion of Williams’s reflections on philosophical style, see Krishnan and Queloz (2022) and Babiotti (2020).

his principle of tolerance towards rival frameworks, and to some extent also to the later Wittgenstein's thesis of the arbitrariness of grammar.²⁰ Within the bounds of what is cognitively and practically feasible for creatures like us, the choice of conceptual framework is rationally underdetermined on this view, and hence arbitrary. It does not matter which framework we choose, as long as the judgements we form within it are true. The choice of framework is understood purely voluntaristically—we can choose one at will. What is not purely a matter of will is which judgements are true and which are false within a given framework. We may freely choose which conceptual framework to bind ourselves by; but, some conceptual framework once chosen, we really are bound by it and by the brute factual recalcitrance of how things are.

On this kind of holistic non-foundationalism, what gives individual concepts their authority is the fact that they are part of a coherent conceptual framework. In the fishing net analogy: what matters is not which weave patterns the net employs, but whether the net allows us to catch fish; and that means that whatever weave patterns we end up going for, the main thing is that the net should not have holes in it. Carnapian frameworks, Quinean webs, or Davidsonian conceptual schemes can all make it tempting to operate with something like this picture in the background. It suggests that a conceptual framework will draw whatever claim to authority it possesses primarily from the extent to which it combines compatible and interlocking concepts into a coherent whole, i.e. a conceptual structure that is free of intra- and interconceptual tensions (I elaborate on the

²⁰ On Carnap's principle of tolerance and his voluntarism, see George (2012), Steinberger (2016), Carus (2017), and Leitgeb and Carus (2021); on Wittgenstein's thesis of the arbitrariness of grammar, see Glock (1996, 'arbitrariness of grammar'), Forster (2004, 67–8), and Kusch (2015). In certain passages, however, Carnap sounds less voluntaristic than he is standardly portrayed as being: 'To be sure, we have to face at this point an important question; but it is a practical, not a theoretical question; it is the question of whether or not to accept the new linguistic forms. The acceptance cannot be judged as being either true or false because it is not an assertion. It can only be judged as being more or less expedient, fruitful, conducive to the aim for which the language is intended. Judgments of this kind supply the motivation for the decision of accepting or rejecting the kind of entities' (1947, 214). Here Carnap seems to envisage something like the goal-based appraisal I consider in Chapter 5; see Reck (2012, forthcoming). In several passages, Wittgenstein also seems to acknowledge a sense in which what concepts we use is not arbitrary. See, e.g.: 'Compare a concept with a style of painting. For is even our style of painting arbitrary? Can we choose one at pleasure? (The Egyptian, for instance.) Or is it just a matter of pretty and ugly?' (2009, II, §367).

notion of a conceptual tension in the next chapter).

The problem with holistic non-foundationalism that emphasizes the arbitrariness and voluntaristic character of framework choice, however, is that it still leaves the particular framework one uses looking rationally contingent: it fails to give us reasons to prefer one framework over another.

Of course, holism that emphasizes the arbitrariness and voluntaristic character of framework choice precisely aims to *deflate* that worry, assuring us that any easily usable framework is as good as any other, and that the questions worth worrying about lie *downstream* of framework adoption, where we worry whether a particular measurement is correct, or a particular thought true. As Wittgenstein replies to the person wondering whether nature really has nothing to say in determining our conceptual scheme: we do ‘run up against existence and non-existence somewhere’, but ‘that means against facts, not concepts’ (1981, 364), and we can only grasp facts downstream of having adopted a conceptual scheme. This invites us to think of conceptual choices in analogy to choices of units of measurement, as Wittgenstein tends to do by comparing ‘rules of grammar’ or ‘laws of inference’ to ‘rules of measurement’.²¹ We might call this the *metrological* conception of conceptual ethics (after *metron*, Greek for ‘measure’). On the metrological conception of conceptual ethics, conceptual frameworks may subtly vary in their usability, but the main difference remains that between having *some* framework and having none.

Once one extends one’s gaze to include thick normative concepts, however, anxiety over whether one has been initiated into the wrong framework reappears: each of these thick concepts contributes, in its own small way, to giving a social world a certain structure and feel, and thereby subtly alters the dynamics and perhaps eventually the trajectory of that social world. In view of this fact, the contention that one framework is as good as another loses its plausibility. More needs to be said for holism to avoid collapsing into ironism at the level of frameworks.

To this end, holistic non-foundationalism requires a more explicitly functionalist underpinning: a society’s conceptual repertoire should be pictured as a harmoniously interlocking whole that is authoritative *for* that society because it has organically grown

²¹ See Wittgenstein (1974, 185; 1978, I, §118). See Kusch (2015) for an overview and discussion of Wittgenstein’s metrological analogies.

out of its particular way of life, and has over time become adapted to that way of life.²²

This elaboration of holism, associated in twentieth-century analytic philosophy with Wittgenstein's notion of a 'form of life', has given a renewed impetus to functionalist holism in anthropology and to communitarianism in social and political philosophy. Not only do a society's concepts harmoniously collaborate in practice, constituting a well-coordinated cultural formation in which each concept performs a function within the whole; the whole that is formed thereby is also uniquely adapted to the character of the society that lives under it, giving them second-order reasons to reason in these terms.

Yet this functionalism creates another problem for holistic non-foundationalism. While it rids itself of the hankering after timeless rational foundations and avoids ironism's indiscriminate disengagement from any and all concepts, it risks doing so at the price of embracing whatever conceptual web we happen to have grown into—an attitude that is not exactly *indiscriminate*, since it remains tightly focused on one's own conceptual apparatus and is withheld towards other ways thinking, but that remains *undiscriminating*, in that it does not draw any finer distinction within that apparatus between better or worse concepts. This is because functionalist holism discourages one from questioning the authority of individual concepts. It suggests, rather, that a form of life must be accepted as authoritative *in toto* or rejected *in toto*. Displace anything, and you risk unsettling a fine balance struck over the ages.

Instead of indiscriminate disengagement from the concepts we find within our form of life, holism thus encourages undiscriminating acceptance of them—in line with Wittgenstein's notorious dictum that 'philosophy leaves everything as it is' (2009, §124).²³ The result is a picture of conceptual authority which revives Right Hegelian conservatism in the guise of a 'Right Wittgensteinianism'.²⁴ In a less pronounced form,

²² See, e.g., Winch (1958), Taylor (1985, 1989), and MacIntyre (1978, 1988, 2007). J. L. Austin's (1961, 130; 1962, 62–64) view arguably also embodies this sort of functionalist holism.

²³ There is, however, room for doubt whether Wittgenstein himself was committed to such a picture across the board—see Williams (2019) as well as Queloz and Cueni (2021); approaches that put a broadly Wittgensteinian picture of our conceptual apparatus to radically critical use include Pleasants (1999, 2002), Celikates (2015) and Jaeggi (2016).

²⁴ See Williams (2005i, 33; 2021, 278) for this use of the label 'Right Wittgensteinianism'. Bloor (1992) also distinguishes between Left and Right Wittgensteinians in order to distinguish 'more historical, social,

something like this holism continues to animate communitarian views in social and political philosophy.²⁵ These invoke, at least as an ideal, homogeneous and harmonious Herderian communities that are tightly integrated by shared thick concepts and traditions. The more coherent and harmoniously integrated with communal practices a conceptual framework is, the harder it becomes to make sense of revisions within it as endogenous and reason-driven (this is the force of Umberto Eco's remark that a worldview can make sense of anything except a different worldview).²⁶

To a lesser extent, a similarly functionalist holism also informs ordinary language philosophy in its Austinian manifestation. For J. L. Austin, we ought at least initially to approach our inherited concepts on the assumption that the concepts and distinctions we find crystallized in 'common sense' embody, *pace* Bertrand Russell, not the metaphysics of the Stone Age, but 'the inherited experience and acumen of many generations of men' (1961, 133).

Austin acknowledges 'in principle' that ordinary language can 'everywhere be supplemented and improved upon and superseded': it may not yet have digested the lessons to be learned from modern microscopes and other expansions of the human perceptual range, for instance, and 'superstition and error and fantasy of all kinds do become incorporated in ordinary language' (1961, 133). Consequently, he grants that 'we may wish to tidy the situation up a bit, revise the map here and there, draw the boundaries and distinctions rather differently' (1962, 63).

'But still', Austin warns,

it is advisable always to bear in mind (a) that the distinctions embodied in our vast and, for the most part, relatively ancient stock of ordinary words are neither few nor always very obvious, and almost never just arbitrary; (b) that in any case, before indulging in any tampering on our own account, we need to find out what it is that we have to deal with; and (c) that tampering with words in what we take to be one little corner of the field is always

and materialist-scientific' (281) interpretations of Wittgenstein from interpretations that are critical of sociological approaches, but Bloor's distinction does not line up with Williams's. A variety of scholars have focused on the structurally conservative aspect of Wittgenstein's thought (Bloor 1983, 1997, 2000; Norris 2009; Nyíri 1976, 1982; Plotica 2015; Rorty 1983; 1989, 58-60; Temelini 2015).

²⁵ See Taylor (1985, 1989), MacIntyre (1978, 1988, 2007), and Sandel (1981, 1996).

²⁶ See Eco (1984, 12).

liable to have unforeseen repercussions in the adjoining territory. Tampering, in fact, is not so easy as is often supposed, is not justified or needed so often as is often supposed, and is often thought to be necessary just because what we've got already has been misrepresented. (1962, 63)

One can ask how conservative this benevolent patience with inherited distinctions makes philosophy—whether it amounts to structural conservatism in practice if not in principle, and whether those distinctions would have been there to be inherited in the first place if people had always been so patient with the concepts they inherited.²⁷

Yet Austin's picture is clearly one on which there is far more latent but coordinated functionality in our inherited distinctions than meets the eye, and these distinctions functionally hang together and cooperate in complex and ill-understood ways. This holism renders it difficult to find, and to be confident of genuinely having found, critical leverage within our conceptual apparatus: it conjures up a serious risk that any putative amelioration will end up being a *Verschlimmbesserung*—an improvement for the worse. In all of these ways, holism encourages concept-users to embrace the concepts they inherited.

3.5 The Kaleidoscopic Picture

It thus seems that we face an unsavoury choice between foundationalism, ironism, and holism: the first no longer makes sense to us when generalized to thick normative concepts; the second encourages indiscriminate disengagement from our concepts; and the third undiscriminating acceptance of them.

To escape this trilemma and find more critical leverage with which to discriminate between concepts that merit confidence and concepts that do not, the first step is to question the holistic picture of our conceptual apparatus as something harmonious, largely tensionless, and inherently static. By ‘picture’, I do not mean the total empirical description one would eventually arrive at after an exhaustive investigation of one’s conceptual apparatus; I mean the *working picture* against the backdrop of which one embarks on such an investigation: the picture expressing one’s default expectations

²⁷ Two charges that Williams (2014d) presses against Austin in a review of his posthumously published essays.

concerning what shape one's conceptual apparatus is likely to take. What kind of working picture one approaches one's concepts with affects what one looks for, what one is primed to see and find salient, what initial stance one takes towards one's concepts, and which questions one is disposed to raise.

A different working picture with which one might approach one's concepts is the negation of the holistic one: it is the picture on which our concepts, far from harmoniously interlocking in a finely calibrated whole, are a historically accumulated jumble—the multifarious outgrowth of diverging, competing, and repeatedly redirected concerns, brimming with intra- and interconceptual tension. While the holistic picture sees the demands of one concept end where those of another concept begin, like cleanly interlocking panels in a stained-glass window, this picture of our concepts—call it the *kaleidoscopic picture*—assumes a motley of overlapping and competing demands. There is no presumption that all one's concepts harmoniously work together, or for one's benefit. Instead, the default stance with which one approaches one's inherited concepts is a sceptical stance. This is also the working picture that Nietzsche recommends:

Hitherto, one generally trusted one's concepts as if they were a wonderful *dowry* from some sort of wonderland: but they are, after all, the legacy of our most distant and most stupid as well as of our most intelligent ancestors. ... What is needed to begin with [*zunächst*] is absolute scepticism towards all inherited concepts. (85:34[195])

The kaleidoscopic picture invites us to approach our concepts not holistically, but on a case-by-case basis, with an open mind as to their merit—one might say that it *encourages* the authority question. Instead of leading one to expect our conceptual apparatus to be mostly tensionless, moreover, it feeds the expectation that it is replete with tensions of all kinds. And instead of presenting that apparatus as a well-calibrated adaption to a certain form of life, it reminds us that a concept might equally turn out to be the legacy 'of our most distant and most stupid' ancestors, and now be of questionable merit, if it ever had any.

On this picture, our conceptual apparatus is expected to harbour great critical potential: extrapolations of some parts of our conceptual apparatus can be condemned, revised, or rejected in light of extrapolations of other parts. The result is a version of the familiar Neurathian strategy of repairing the raft that is our conceptual apparatus not in drydock, but out on the open waters, mending one plank while resting on the others.

Insofar as our conceptual apparatus offers the inherent critical leverage that the kaleidoscopic picture invites us to seek out in it, the Neurathian strategy is capable of finer discriminations within our conceptual apparatus than either the Carnapian or the Wittgensteinian holistic strategies (perhaps uncoincidentally, Neurath was also the most politically engaged member of the Vienna Circle).²⁸ This inherent critical leverage in turn invites one to think of our conceptual apparatus not as inherently static, but as inherently dynamic: a structure whose change over time becomes intelligible as endogenous and reason-driven rather than as an exogenous and merely causal imposition from reality.

This is a case where knowledge of the history of something can inform our expectations about it and structure our interpretation of it. If one learned that a text was not carefully masterminded by a single author, but in fact constitutes a collage of snippets assembled and rewritten by several authors over many generations, one would not be surprised to find the text full of tensions and incoherences.²⁹

Analogously, knowledge of the history that produced our concepts can inform our expectations as to how coherently they are likely to fit together, and thereby give us grounds for favouring one working picture over another. It would be nothing short of astonishing if the concepts we inherited from centuries of history—shaped as they are by countless historical contingencies, appropriated, extended, transformed, amalgamated, and repurposed by countless different factions many times over—should have ended up forming, of all things, a harmonious whole.

Even if we accept that selective pressures of one sort or another were at work in the history of our concepts, this only yields the conclusion that each concept might individually have been subject to certain selective pressures—there might be selection of some sort at the level of the individual concept, resulting in the concept becoming adapted to deliver whatever effects are being selected for. Without a reason to think that there was strong selection *for coherence*, however, it does nothing to support the assumption of coherence; and why would evolution, cultural or biological, have selected above all for coherence, of all things? Given what we know about the variegated history

²⁸ See Edmonds (2020, ch. 12). For a discussion of the many affinities between Otto Neurath's thought and present-day ameliorative conceptual engineering, see Yap (2022).

²⁹ For a detailed argument to the effect that knowledge of the history of something can yield an interpretative structure guiding our interpretation of it, see Prescott-Couch (2015, manuscript).

of our concepts, then, the kaleidoscopic picture should be the default picture as we approach the concepts that have come out of that history.

Another way to put the point is this: the holistic picture needs to be supported by an account of how the coherence it assumes *got there*, and that account needs to be less controversial, or at least differently controversial, than the picture it is meant to support. The kaleidoscopic picture, by contrast, does not call for a supporting account in this way. It is not a picture of a different controversial assumption, but a picture of the absence of such an assumption.

Adopting a kaleidoscopic picture as our working model of the conceptual apparatus marks the first step towards escaping the trilemma of foundationalism, ironism, and holism. When developed against such a backdrop, a non-foundationalist approach to concept appraisal can count on there being plenty of critical leverage in our conceptual apparatus already in virtue of the intra- and interconceptual tensions in it. If extrapolations of some parts of our conceptual apparatus can be condemned, revised, or rejected in light of extrapolations of other parts, it becomes intelligible how there can be endogenous, reason-driven conceptual change.

3.6 Leveraging Local Needs

But the non-foundationalist approach to conceptual authority I propose to develop here does not simply draw its critical leverage from the tensions to be found in our conceptual apparatus. It finds a leverage point that is more extraneous than that, though less extraneous than timeless rational foundations. As we shall see in more detail in Chapter 6, it finds that leverage point in concept-users' *conceptual needs*—the needs they have for certain concepts as a result of their concerns, capacities, and circumstances. This is the second step out of the trilemma. The approach to conceptual authority by which I propose to escape the trilemma is thus a *needs-based* non-foundationalism: an appraisal of the concepts we have according to the concepts we need.

At first pass, the proposal that we should appraise our concepts on the basis of our needs as construed through those very concepts appears to make such a needs-based non-foundationalism (a) circular, and (b) sensitive to local circumstances. In the remainder of this chapter, I argue that any appearance of problematic circularity evaporates under analysis, and that sensitivity to local circumstances is a feature rather

than a flaw of the approach: it manages to draw additional critical leverage from local needs.

Take circularity first. Tristram McPherson and David Plunkett (2021) have formulated a ‘vindictory circularity challenge’ for conceptual ethics, arguing that since any evaluation of our concepts itself has to draw on evaluative concepts, reflection that succeeds in vindicating concepts is likely to display the circularity involved in evaluating a standard by itself. Yet, as Nietzsche already observed: ‘*Ein Werkzeug kann nicht seine eigene Tauglichkeit kritisiren*’ (85:2[132])—a tool cannot critique its own adequacy. An important challenge for conceptual ethics therefore lies in explaining why this circularity does not vitiate the enterprise of conceptual ethics.

If that enterprise is described as appraising our concepts in terms of those same concepts, it indeed looks circular. But that is still an insufficiently determinate description that requires further disambiguation and concretization before a verdict can be reached one way or another.

That description that could, for instance, be taken to designate the enterprise of turning each individual concept’s evaluative force back on itself, by asking whether the concept *F* was itself *F*—whether the concept *precise* is itself precise, in McPherson and Plunkett’s example.³⁰ Where this strategy of conceptual auto-evaluation is intelligible at all, however, it is nothing like checking a ruler against itself. It seems well worth asking whether the concept *useful* is itself useful, the concept *helpful* itself helpful, or the concept *life-enhancing* itself life-enhancing. These concepts are not trivially self-validating in the way that checking a ruler against itself would be. For one thing, their content and evaluative force only become fully determinate once one specifies the contextual and conceptual background against which they are deployed (useful to what end? Helpful for whom? What counts as an enhancement?). But the main difficulty with this strategy of conceptual auto-evaluation is that it tends to produce nonsense when generalized: it simply makes no sense to ask whether the concept *ultraviolet* is itself ultraviolet, the concept *happy* itself happy, the concept *ambidextrous* itself ambidextrous, and so on for the overwhelming majority of concepts.

³⁰ An interpretation that McPherson and Plunkett (2021, 214–5) also consider before moving on to more holistic interpretations; as they themselves note, they are best read as raising an entire family of related circularity challenges.

Needs-based appraisal does not involve vindicating a concept *F* in terms of that same concept *F*, however. Rather, the idea is to disengage oneself from the concept *F* and assess its authority by examining the conceptual needs one has as a result of one's concerns, capacities, and circumstances. A concept is thus to be appraised by the lights of an array of items that are *not* concepts. It is merely that in individuating these items, one is bound to make engaged use of the rest of one's concepts, or at least of some of them.

While I shall argue in later chapters that we can have conceptual needs unwillingly and unwittingly, it is of course trivially true that what conceptual needs we can conceptualize ourselves as having is partly a function of our concepts. In that sense, the leverage point provided by our needs for certain concepts is not entirely independent of our conceptual apparatus: in the terminology developed in Chapter 1, we might say that our conceptual needs are hermeneutically, though not necessarily logically or ontologically, dependent on our conceptual apparatus: we can only *make sense* of them as needs of ours by virtue of our concepts.

But we are not searching—incoherently—for a leverage point that is extraneous to our concepts in the sense that its reflective recognition does not draw on our conceptual resources at all. We are bound to draw on our concepts in bringing *any* standard to bear on our reflection, even if it is the most external Archimedean point. As John McDowell also stresses, ‘one can reflect only from the midst of the way of thinking one is reflecting about’ (1996, 81). Yet the fact that our needs only come within our conscious purview by dint of the concepts through which we conceptualize them is no bar to seeing those needs as extraneous to our concepts—just as the fact that one only recognizes *fouls* as such if one has the concept thereof does not turn tackles into concepts.

Such a needs-based approach can also avoid the second, more holistic form of circularity that McPherson and Plunkett (2021, 214) are worried about, where a set of concepts mutually support each other in a way that systematically forecloses the possibility of critique. For one thing, this is a problem one would expect to be pervasive only on the holistic picture of our conceptual apparatus; on the kaleidoscopic picture, we do not expect our concepts to prove mutually vindicating—certainly not across our conceptual apparatus. On the contrary, we expect there to be no shortage of critical movement as tensions are alleviated and new ones created.

Furthermore, combining a kaleidoscopic picture of our first-order reasons with an

approach that reaches beyond those first-order reasons to find a basis for critique in conceptual needs promises to give us a critical grip even on mutually vindicating sets of concepts. This leverage point in needs is independent enough to render intelligible the possibility that even a community of concept-users whose conceptual apparatus formed a harmonious and tensionless whole could find reasons to become radically dissatisfied with large swathes of that apparatus. It might no longer be what they need. Making concepts answerable to our needs thus allows in principle for profoundly radical critique.

My suggestion is not that we should assess the authority of all our concepts at once, moreover. Doing so would once again introduce a problematic form of circularity, since it would mean that at least some of our concepts would have to figure at the same time among the objects of appraisal and among the concepts used in an engaged way to appraise them. Rather, the idea is to operate piecemeal and disengage from *one* concept, or one connected *set* of concepts, to assess its authority based on our conceptual needs as construed through the engaged use of the *rest* of our concepts. As a result, the appearance of circularity evaporates under analysis. The concept appraised does not itself contribute to defining the standard of its own appraisal, and no concept is safe from revision: the insistence that we cannot question everything at once is entirely compatible with the idea that everything is open to question.

There is, however, one respect in which needs-based appraisal retains something of the self-referentiality which marks off non-foundationalism from foundationalism. If our conceptual needs are, as I shall argue, the product of how our concerns interact with our capacities and circumstances, then our assessment of the authority of a given concept will depend notably on what concepts figure among ‘the rest of our concepts’. Not only what we take our concerns, capacities, and circumstances to be, but what they in fact *are* depends on what concepts structure our social world. This is most obviously true of the concerns we pursue: concepts can instil new concerns—such as the Enlightenment concern for autonomy, or the Romantic concern for authenticity—that we would not have without a host of concepts putting the relevant considerations, desires, and aspirations within our cognitive grasp. Here we really have not just hermeneutic dependence, but ontological and logical dependence: the concerns for autonomy and authenticity only exist in virtue of the concepts we possess, which also makes the reasons given to us by these concerns dependent on those concepts. What we care about and

want is a function of the concepts we possess. A more prosaic example is provided by Henry Ford, first mass-producer of cars, who quipped that if he had asked people what they wanted, they would have said: ‘faster horses’.³¹

Similar dependences on our concepts are exhibited by our capacities and circumstances. These are so deeply affected by the conceptual architecture we inhabit that no accurate construal of these capacities and circumstances for the purposes of appraising one of our concepts could reasonably hope to avoid reflecting the influence of the rest of our conceptual architecture. Consequently, in inquiring whether we need anything like the concept *F*, and what the concept we need looks like, we will draw on the rest of our concepts, and this means that what conceptual needs we can discover that we have is not completely independent of the concepts we possess already.

The only way to escape this dependence would be to view all our concepts as answering only to needs we have anyway, i.e. antecedently and independently of any of our concepts. But this would be to take a very narrow view of our needs, akin to that adopted by the evolutionary psychologist wondering how the human artefacts exhibited at MoMa might be conducive to survival and reproduction.³² The advent of new concepts can lead to the creation of new and sociohistorically local needs. If we allow these local needs to inform concept appraisal, our concepts will be made to answer to needs we would not have without them. But there is no problematic circularity involved in this as long as the concepts needed and the concepts involved in engendering and recognizing the needs are distinct concepts.

This brings us to the sensitivity to local circumstances which I suggested was a feature of the approach. The non-foundationalist conviction that the present book is guided by is that we want *our* concepts to help *us* to live, and they can do this, most notably, by helping us to meet *our needs*. These needs are clearly not entirely, or even mainly, those of disembodied intelligences floating free of localized historical developments. We need concepts that are rooted in our local perspective, a perspective that is reflective not only of our humanity, but also of our distinctive identity, projects, and commitments. If the

³¹ In the literature on design thinking and innovation, this limitation that our extant concepts impose on our driving concerns has accordingly been called ‘the faster horses trap’ (Gordon, Rohrbeck, and Schwarz 2019).

³² As Miller (2000) does, for example.

authority of a concept is tethered to the local concerns, capacities, and circumstances creating a need for it, however, the variability of these factors across people, places, and periods means that there is no one set of concepts that is universally and eternally best. Much as there is no absolutely and definitively best set of tools for an artist, because the tools have to suit the artist and the kind of artistic vision she is pursuing, what concepts one needs becomes a function of one's concerns, capacities, and circumstances, and varies with them. Again, the connection between concepts and concerns goes both ways: just as the advent of new tools has led artists to strive for different forms of art, the development of new concepts can instigate the development of new concerns and alter our capacities and circumstances. There are many needs we would never have acquired were it not for the emergence of certain concepts. But when we assess the authority of a particular concept, it is, in the first instance, our conceptual needs that we should look to. How well the concept meets those needs will determine how authoritative it is, and whether a rival concept is more authoritative because better tailored to those needs.

Some of our needs will be best met by minimally perspectival concepts that we might expect even very differently situated intelligences to converge on; but the totality of our conceptual apparatus should answer to the totality of our needs, and it would be an impoverished human life that had no more local needs besides those near-universal ones. A conceptual apparatus fit to meet our various needs cannot be limited to the bloodless abstractions we might expect even alien thought to grasp. It must include thickly perspectival concepts that will reflect our particular historical, cultural, and social situation. Precisely because we are not, in Williams's phrase, 'unencumbered intelligences selecting in principle among all possible outlooks' (2006g, 193), we need to acknowledge that concepts are not just authoritative from a perspective that happens to be ours. Rather, that perspective is the source of their authority.

The pressure on concepts to reflect local peculiarities might be thought to be implicitly recognized in the widespread—and widely abused—fascination exerted by concepts expressible only in hard-to-translate words from other cultures.³³ What fuels this fascination is the hope that the most distinctive parts of a culture's vocabulary might

³³ The fascination is abused, for example, when utterly banal advice is marketed by dignifying it with some allegedly hard-to-translate term from another culture and dressing it up as rarefied ancient wisdom (often at the expense of translational accuracy).

act as a guide to the peculiarities of its circumstances and its most distinctive concerns. That is not always true, since, in inferring from words and concepts to concerns, one is liable to mistake a mere contingency of linguistic evolution for a marker of deep cultural difference. But when moving in the opposite direction, from concerns to the concepts they call for, there is a robust connection that can more reliably guide our understanding of what concepts we have most reason to use: the peculiarities of our situation and the distinctive concerns we pursue in it do make it worthwhile for us to have certain concepts rather than others.

To think that our concepts should answer to something as local and variable as human needs is to adopt a contingent standard—a standard that is not only properly expressive of our humanity (non-foundationalism has been aptly called a ‘methodological humanism’),³⁴ but also the product of even more local and contingent sociohistorical forces. Such a standard will not yield many reasons for concept use that would be recognizable to any rational agent. But it will yield reasons for those who share the relevant concern and the kind of situation in which it is pursued. This is not to settle for second-best after relinquishing all hope of finding timeless rational foundations. The question should not be whether *anyone* has reason to prefer the concepts we have over alternatives, but whether *we* do.

In its willingness to tailor concepts to genuinely local needs, the present approach goes decidedly beyond philosophical approaches that aim to separate out the concepts that human beings *necessarily* have from those that they only *contingently* have. Philip Pettit, Miranda Fricker, and Robert Smithson, for example, have recently articulated philosophical research programmes aiming to map out the constraints on our conceptual scheme that arise from human beings’ *most constant* concerns: concerns that tend to be at work in any human community, no matter its location in space and time.³⁵ Of course, these constant concerns can only engender necessities that are themselves contingent upon certain highly general facts about human nature or social arrangements. But, as Jonathan Rée observes: ‘Contingencies can last a very long time. Our preoccupations with love and death may not be absolute necessities, but they are not a passing fad either, and it is a safe bet that they will last as long as we do’ (1998, 11).

³⁴ See Lauener (2001).

³⁵ See Fricker (2020a), Pettit (2018, forthcoming), and Smithson (2020).

Accordingly, the research programme of ‘conceptual cartography’, as Smithson calls it, seek to isolate those features of our conceptual scheme that are necessary for creatures with our basic nature, in the sense that ‘we cannot imagine humans accomplishing their basic projects without having a conceptual scheme with these features’, from features that are contingent, in the sense that ‘we can imagine communities effectively using a somewhat different conceptual scheme’ (2020, 1).³⁶ Transcendental arguments as advanced in different forms by Kant, Donald Davidson, and P. F. Strawson provide one kind of template for this.³⁷ The concepts that are necessary form what Strawson calls that ‘central core of human thinking which has no history—or none recorded in histories of thought’ (1959, 10).

One reason to want to understand ‘to what extent features of our actual practice are necessary, and to what extent they are contingent’, as Fricker puts it, is that this ‘will in turn explain how some kinds of criticism of our practice are worth making, and how some are senseless’ (1998, 165). Furthermore, identifying the features of our conceptual scheme that grow out of absolutely basic and anthropologically universal concerns promises to allow us to ‘specify the limiting conditions on our exercise of ethical freedom’ (2020a, 931), Fricker argues: while ‘we are, in a far-reaching sense, *ethically free*’ in the sense that ‘we are substantively free to set our own ends, and thereby generate our own values and correlative practical reasons’ (2020a, 921), we are ethically free only *within certain limits*—there are concepts that human beings cannot do without for long. Of course, it is a notorious truth that human beings are capable of destroying even what they absolutely need. But just because of this, it might be thought that identifying the elements of our conceptual scheme that we cannot do without is important, as it can help us to check dangerously destructive impulses.

The exclusive focus of such approaches on our most constant and most universal needs expresses an evaluative assumption that is questionable for the purposes of

³⁶ Though Smithson goes on to distinguish different species of necessity, one of which—‘pragmatic necessity’—he relativizes to sociohistorically situated communities. A feature F is then pragmatically necessary for a community, according to Smithson, iff ‘all best suited languages for that community contain F’ (2020, 12). Fricker also distinguishes species of necessity, including the practically necessary and the humanly necessary that arises out of ‘human emotional nature’ (2019, 245).

³⁷ See Smithson (2020, 17–18).

conceptual ethics, however, namely that what answers to universal needs is *more important* than what reflects more local factors. This evaluative assumption comes out well in Pettit's description of the research programme he envisions:

Which are the more or less passing ephemera and which the phenomena that are deeply embedded in the society? Which are more or less incidental or contingent features and which are features apt to last? There is an interesting research programme suggested by such questions. It would take any society or culture or institution and, reviewing the data on various traits displayed by the entity in question, would seek to separate out the dross from the gold. It would try to identify and put aside the features that may be expected to come and go. And it would seek to catalogue the more or less necessary features that the society or culture or institution displays. It would give us a usefully predictive stance on the society, providing us with grounds for thinking that such and such features are likely to stay, such and such other features likely to disappear. (1996, 299–300)

The description of what is held in place by universal human needs as 'gold' and of the rest as 'dross' makes the evaluative hierarchy underlying this sort of approach fully explicit. In fairness to Pettit, it need mean no more than that the former is gold *for the purpose* of predicting what is likely to stay in place. But that still expresses the evaluative assumption that this should be the purpose of the exercise.

From the perspective of conceptual ethics, there are good reasons to question these evaluative assumptions: it is not obvious that what conceptual ethicists should primarily care about is the distinction between what is pinned in place by immutable needs and what is not, and even less obvious that the former should be elevated above the latter. In particular, there is no reason to think that the distinction between authoritative and unauthoritative concepts aligns with the distinction between what is pinned in place by immutable needs and what is not. Concepts might answer only to sociohistorically local needs, and yet be no less urgently needed for that.

What is more, these approaches focusing on the limiting constraints imposed on our conceptual freedom by immutable necessities are simply silent about which concepts we have reason to prefer within the bounds delimited by those universal necessities. It may well be that what answers to universal human needs is gold, but it does not follow that everything else is dross—there are other precious materials. Marking out universally necessary concepts and delineating the bounds of our conceptual freedom is a start. But

a truly discriminating and widely useful approach to concept appraisal should be sensitive to further distinctions between concepts, and allow us to identify reasons for reasons *within* the space of our conceptual freedom. It should not just demarcate the outer limits of conceptual change, but guide it.

The key to identifying reasons for reasons even within the space of our conceptual freedom is to recognize that which concepts are ‘necessary for us’ is itself a function of *who we are*, and therefore notably a function of more local historical and cultural forces: some concepts may be necessary *for us* without being necessary for every human community, because we have needs and concerns that others did not have, or did not have under the same circumstances. The fact that these concepts are only *locally and contingently* necessary, in that they are necessary in virtue of concerns and circumstances that are to varying degrees local and contingent, does not detract from their importance, because it does not detract from the urgency of their necessity.

Hence, while the approaches of Pettit, Fricker, or Smithson focus on the broad-meshed anthropological necessities that constrain our otherwise free and contingent conceptual choices, I aim to identify, in our more local needs and concerns, a basis from which to discriminate between better or worse concepts even within the realm of the contingent: there are concepts that *we* have reason to use in virtue of our local needs, though not every human community does. These reasons do not merely act as negative constraints on an otherwise rationally undetermined horizon of conceptual freedom, but offer positive guidance and criteria by which to evaluate different ways of thinking within the horizon of what is possible for creatures like us.

Lest this view of conceptual authority as a function of the needs created by our concerns, capacities, and circumstances seem obvious, however, it is worth pointing out that there is another, far more influential way out of the trilemma. This rival non-foundationalist approach to conceptual authority treats it not as something that is conferred upon concepts by the needs of concept-users, but as something inherent in features of the concepts themselves—in particular, in their theoretical virtues, such as their precision, determinacy, consistency, or coherence. On this view, which also operates against a picture of our conceptual apparatus as ridden with intra- and interconceptual tensions and defective in various ways, the *tidier* way of thinking is the more authoritative one, and philosophers’ claim to attention derives notably from their skill in tidying up

our slovenly thought.

It is to understanding the motivations and gaps in this rival non-foundationalist approach that we now turn—the approach that seeks conceptual authority through the tidy-minded pursuit of theoretical virtue

CHAPTER FOUR

Tidy-Mindedness

The non-foundationalist answer to the authority question suggested by some of the recent literature on conceptual engineering is that we can criticize and improve our inherited conceptual apparatus, however parochial, by tidying it up.¹ As advocates of conceptual engineering emphasize, our inherited concepts tend to be in various ways messy and defective: unclear, imprecise, vague, gerrymandered, inconsistent, or incoherently related to other concepts. By engineering our concepts to be more precisely defined, more consistent, or more coherently related to other concepts, we can fix those defects, and arrive at a set of concepts that has a better claim to being authoritative than the concepts we started out from. In contrast to what the traditional opposition of ‘foundationalism’ to ‘coherentism’ in theories of truth and justification would suggest, however, this non-foundationalist ideal does not just pull towards coherence.² It pulls towards the entire gamut of virtues paradigmatically exemplified by an axiomatized theory.

Philosophers’ pursuit of tidy ways of thinking—their *tidy-mindedness*—has a long history: ‘philosophers have always aimed at cleaning up the litter’, Williams James notes—they have aimed to replace ‘the first sensible tangle’ with conceptions that are ‘intellectually neat’, ‘orderly’, and ‘always aesthetically pure and definite’ (1975–88, vol. 9,

¹ See, e.g., Brun (2016, 2020), Cappelen (2018, ch. 2), Simion and Kelp (2020), Eklund (2002, 2019, 2021), Scharp (2013, 2020, 2021), Greenough (2020), and Dutilh Novaes (2020a, b; 2017).

² See Olsson (2017) and Hasan and Fumerton (2018) for overviews.

26). This tidy-mindedness is manifest not only in how philosophers treat their chosen subject matters, but also in how they tend to gravitate towards subject matters that *admit* of tidy treatment.

Yet, despite James's reference to the aesthetic dimension of philosophers' tidy-mindedness, the pursuit of tidy ways of thinking is not just an aesthetic quirk. It is more charitably understood as an attempt to pursue more authoritative concepts, and to guard against slovenly thinking and the risks it brings by making our concepts more theoretically virtuous. Because 'the slovenliness of our language makes it easier for us to have foolish thoughts', language needs 'fixing' (2008, 270), as George Orwell already put it in a phrase echoed by the title of Herman Cappelen's seminal monograph on conceptual engineering, *Fixing Language* (2018).

On a tidy-minded approach to concept appraisal, a concept, or a constellation of concepts, counts as authoritative to the extent that it realizes a catalogue of *theoretical virtues*, i.e. the virtues that are prototypically those of a neatly axiomatized theory: clarity, determinacy, precision, fruitfulness, consistency, coherence, etc. We can discriminate between more or less authoritative concepts according to the degree to which they exhibit theoretical virtues or theoretical vices.³

Importantly, this is not a foundationalist standard. It is an ideal that drives us towards a tidied-up version of whatever conceptual apparatus we start out from, not *out* of those inherited ways of thinking and towards the set of concepts that is absolutely and definitively best. It is a purely formal ideal, indifferent to the substantive content of concepts, that can be satisfied in a plurality of ways. We then seek not the uniquely best mind, but a tidy mind, stocked only with definite and neatly arranged concepts. (It would take a further, extremely strong assumption to the effect that the formal pursuit of theoretical virtues alone would already be enough to ensure converge on a single set of concepts, whatever one's starting point, to turn this formal and non-foundationalist pursuit into a substantial and foundationalist one.)

My aim in this chapter is to assess the merits of this tidy-minded approach to concept appraisal as a general solution to the problem of conceptual authority, and in the process to clarify various kinds of theoretical vices that concepts might display.

³ As recently argued by Wakil (2021), for example.

4.1 Theoretical Vices in Concepts

Like the needs-based approach, the tidy-minded approach does without controversial assumptions to the effect that our conceptual apparatus forms a coherent whole, or that it is uniquely adapted to the society it has grown out of. On the contrary, the tidy-minded approach conceives of our conceptual apparatus as a jumbled product of history presenting us with many theoretical vices that a tidy-minded view of conceptual authority can fasten on. As Cappelen writes: ‘it’s implausible that a cultural artifact that’s generated in a messy, largely incomprehensible way that’s outside our control should end up producing something we can’t improve on’ (2020, 139).

A concept might, for example, be insufficiently ‘clear and distinct’, in the influential Cartesian phrase, leaving its content hard to make out and ill-demarcated from other things. On the tidy-minded approach, this is a conceptual defect we have to remedy. As Descartes insists in a letter to Mersenne: ‘We have to form distinct ideas of the things we want to judge about’ (1996, 3:272).⁴

In more contemporary idioms, we can distinguish a great many conceptual defects, some of which offer us different ways of spelling out the Cartesian ideal of clarity and distinctness, and some of which extend it to encompass additional theoretical virtues. A concept might suffer from referential indeterminacy, for instance, so that there is no fact of the matter as to what the concept refers to; or the inferential relations a concept stands in might be insufficiently determinate, leaving it unclear what its applicability is entailed or excluded by, and what it itself entails and excludes; or, like the notorious concept *heap*, a concept might be vague, i.e. lack sharp boundaries and give rise to cases in which one does not know whether the concept is applicable or not.

A lesser-known theoretical vice a concept might display is to be *open-textured* or *porous*: vulnerable to the advent of circumstances in which there would no longer be a fact of the matter whether the concept applied or not. Porosity is not the same as vagueness. Friedrich Waismann’s point in introducing the notion of the *Porosität der Begriffe*, the porosity of concepts, was that, although concepts such as *gold* or *mother* might be exactly defined and their applicability under different circumstances fully determinate, the discovery of new elements or biotechnological advances might give rise

⁴ On Descartes’ notions of clarity and distinctness, see Paul (2020).

to novel circumstances that made it unclear whether the concepts were applicable.⁵ The concepts are thus porous in the sense that not ‘every nook and cranny is blocked against entry of doubt’ (Waismann 1945, 123). They are not vague, but vulnerable to *becoming* vague.

A concept might also fail to pick out natural kinds and instead delineate gerrymandered kinds, held together by nothing more than our willingness to treat them as a kind while cutting across clusters of properties that are in fact unified by more than that. Relatedly, a concept might be unfruitful as a result of subsuming under one heading what is better kept distinct, as Carnap thought was the case with the non-scientific concept *fish*, which includes whales and dolphins and thereby allows us to formulate fewer law-like generalizations than the scientific concept *piscis* (Carnap’s term), which excludes whales and dolphins.

Carnap understood the fruitfulness of a concept in terms of its conduciveness to formulating what he called ‘universal statements’, by which he meant ‘empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept’ (1962, 7). But not all domains of inquiry are now thought to fit that description. Many do not aim at either empirical or logical statements, and even if they do, they do not necessarily aim at universal ones. This had led to various proposals for amendments to the notion of fruitfulness, on which a concept counts as fruitful to the extent that it furthers the aims of scientific inquiry,⁶ facilitates the production of new knowledge,⁷ or facilitates progress towards the achievement of theoretical goals,⁸ and unfruitful to the extent that it hinders these things. These amendments make fruitfulness a great deal broader and harder to quantify than Carnap’s conception, which enabled one to simply tally up the number of universal statements containing the concept.⁹ Yet even these amendments leave the notion of fruitfulness narrowly restricted in one significant respect: it remains sensitive only to what a concept does for scientific inquiry, the production of new knowledge, or

⁵ For further discussion and clarification of the concept of open texture, see Vecht (2020).

⁶ See Kitcher (2008).

⁷ See Dutilh Novaes and Reck (2017).

⁸ See Pinder (2022).

⁹ Some of Carnap’s formulations at least suggest this as a criterion for the quantification of fruitfulness; see, e.g., Carnap (1962, 14–15); others suggest a wider and hazier criterion, on which a concept is fruitful to the extent that it ‘leads to more simple and interesting theorems’ (Carnap 1962, 348).

theoretical goals, when—as the next chapter will bring out—much of what concepts do for us lies outside the perspective provided by these aims, and may even work against them.

What all of these theoretical vices in individual concepts certainly do, however, is to give the tidy-minded approach critical leverage over our conceptual apparatus. On the tidy-minded conception of conceptual authority, more authoritative concepts can be arrived at by replacing these defective concepts with more theoretically virtuous alternatives. That approach to conceptual authority has a distinguished history—it is as Cartesian as it is Carnapian.

4.2 Superficial Concepts

Besides these well explored theoretical vices, there is another familiar, but far less theorized complaint one might have about a way thinking, namely that it is *superficial*. Not all the ways in which thought might be said to be superficial invite description as a theoretical vice. But it is a widely shared idea that one virtue of a good scientific theory is its *depth*, and that, correspondingly, superficiality is a vice in a scientific theory. Even theories that enable one to formulate perfectly true judgements articulated in terms that are clear, determinate, non-vacuous, and fruitful may nevertheless feel unsatisfactory due to their superficiality. When focused on individual concepts, this yields the demand that our concepts should be deep rather than superficial, and that superficiality is a theoretical vice in a concept.

This indicates an additional dimension along which one might pursue conceptual authority through theoretical virtue: by replacing superficial concepts with deeper concepts whenever possible. But how exactly should we understand the contrast between depth and superficiality as it applies to individual concepts?

Extrapolating from Michael Strevens's (2008) two-dimensional development of the notion of *depth* in scientific explanation, we can understand a concept's relative superficiality or depth as determined by how it scores along two dimensions: its attention to causal detail on the one hand, and its causal generality on the other.

Along the first dimension, a concept is superficial to the extent that it slights, ignores, or skirts the causal underpinnings of what it picks out. A superficial concept may still pick out something, and may attach a certain significance to it, but it is largely indifferent to—

and hence uninformative about—the causal underpinnings of its extension: it does not reach far down into the physical level at which the ultimate causal details are to be found, or far into the causal history of its extension. That is not to say that the superficial concept is entirely indifferent to causal detail; but it abstracts away from any detailed description of the causal processes that underlie it, operating at a higher and more superficial level of description.

By contrast, a concept will be deep along this first dimension to the extent that its application is sensitive to and informative about causal underpinnings. Concepts that are deep in this way will tend to be epistemically more demanding than superficial ones, requiring one to dig deep into the aetiology of phenomena in order to determine whether or not the concept applies.

Along this dimension of depth as attention to causal detail, the concepts of everyday psychology—*belief*, *desire*, *intention*, etc.—for example, are still relatively superficial: their application is guided by easily observable and publicly accessible patterns of behaviour and speech, and while their application can be sensitive to proximate causes at the most ordinary level of description ('Did the gust of wind make you drop the daisy, or did you do it intentionally?'), they are indifferent to the deeper causal underpinnings of those psychological phenomena. The concepts of cognitive science with which some strive to replace these 'folk' psychological concepts, by contrast, promise to be neurophysiologically deeper—that is part of their attraction.¹⁰ Other examples include moving from superficial concepts of taste to deeper concepts articulating the underlying chemistry (from the concept *salty* to the concept *contains NaCl*; or from the concept of a *flinty* wine to the concept of a wine containing high levels of sulphur dioxide).

Part of the attraction of deeper concepts is that their sensitivity and informativeness regarding causal underpinnings promises to make them more objective. Judgements of ambient temperature, for example, had long been made in terms of comparatively superficial concepts, such as *warm* and *cold*, that were indexed to subjective experience. The deeper and more scientific concept of temperature, which explains these subjective experiences in terms of their objective causal underpinning, makes judgements of temperature more objective, giving people an independent measure against which their subjective experience can be compared. But, as Hasok Chang shows in his detailed

¹⁰ See Stich (1983) and especially Churchland (1986).

account of how the scientific concept of temperature developed, the path to such a deepened concept was a long and winding one: even in the nineteenth century, scientists such as Joseph Fourier still articulated their theories of thermal physics in a ‘macroscopic-phenomenalistic vein’, conceptualizing temperature in terms that remained ‘noncommittal about the ultimate metaphysical nature of heat’ and ‘did not focus on considerations of “deep” causes’ (Chang 2004, 96–97). Once concepts are deepened, however, they provide a lower-level understanding of the causal processes underlying what they refer to, and enable one to articulate commitments about the ultimate metaphysical nature of those processes.¹¹

The second dimension along which concepts can be deep is by cutting through superficial detail to reveal a hidden structure of great explanatory power. This is not depth through careful attention to causal detail, but depth through striking generality: concepts that are deep along this second dimension uncover abstract dynamics that are not tied to a cumbrous set of specific initial conditions, but depend only on the presence of a handful of abstract properties that can be found across a range of different conditions.

By acknowledging this second dimension of depth, one acknowledges that while attention to causal detail may be a large part of what is involved in achieving the theoretical virtue of depth, it is not *all* that is involved, because we clearly also do not want to become bogged down in the minutiae of causal processes—that way, Strevens warns, lies ‘Laplacean blindness’ to the higher-level structures that constitute the furniture of human affairs.¹² We also want our scientific concepts to abstract away from causal factors that only affect how exactly something happens in one particular instance,

¹¹ This conception of what is involved in deepening a concept is meant to parallel Strevens’s (2008, 129–33) account of what is involved in deepening a causal model.

¹² See Strevens (2008, 138–41). The reference is to Pierre-Simon Laplace’s evocative description of what it would mean to achieve maximum depth along the first dimension in a deterministic universe: ‘An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes’ (Laplace 1951, 4). As Elliott Sober (1984, §4.3) and Daniel Dennett (1989, 25) have emphasized, Laplace’s imagined intellect, now often referred to as ‘Laplace’s Demon’, helpfully dramatizes the question of what, if anything, such an intellect would be missing.

and focus on those causal factors that decisively make a difference to *whether* something happens. This second dimension of depth thus involves tracking the select few abstract properties that reveal high-level dynamics of some generality. This is how we acquire the kind of understanding that stands a chance of helpfully *carrying over* to other situations.

As Strevens shows, recognizing this second dimension of depth allows us to make sense of the apparent disdain for causal detail exhibited by many forms of scientific explanation—paradigmatically, by equilibrium explanations of regularities in complex systems, such as R. A. Fisher’s 1930 explanation of the remarkably regularly observed one-to-one sex ratio among sexually reproducing organisms: individuals of the less frequent sex have more reproductive opportunities than individuals of the more frequent sex until this one-to-one ratio is reached. This explains the ratio not by drilling down into the causal details of how it came about in each instance, but by showing, at a more abstract level, that it marks an equilibrium point under negative frequency-dependent selection.¹³

A similar slighting of causal detail can be observed in idealizing explanations, such as the explanation of the cannon ball’s parabolic trajectory that ignores air resistance, or the explanation of the rainbow that falsely assumes raindrops to be perfect spheres: in fact, each cannon ball encounters some air resistance, and each raindrop is slightly deformed by local forces; but these idealizing explanations embody the insight that these causal factors make no difference to the trajectory’s approximate shape or to the appearance of the rainbow. These ways of scientific sense-making are thus not so much uninterested in causal detail as concerned to home in on those causal details that constitute generalizable difference-makers, i.e. factors whose presence or absence is apt to make a difference.

The concern with depth as attention to detail thus has to be balanced against the concern with depth as generality: we want our scientific explanations—and, by extension, the concepts that figure in them—to be as sensitive to detail and informative as possible while remaining as abstract and generalizable as possible.

But, however the balance between these two dimensions of depth is struck, the basic idea remains that depth in either dimension is a theoretical virtue, and superficiality in

¹³ See Strevens (2008, 137). Another example he offers is Ludwig Boltzmann’s explanation of the second law of thermodynamics, whereby the entropy of an isolated system always increases to a maximum equilibrium value.

either dimension a theoretical vice. The tidy-minded approach to conceptual authority, insofar as it proposes to generally model authoritative concepts on those of a good scientific theory, will therefore encourage us to make our concepts as deep as possible, and to treat superficiality as a defect. As we shall see in Chapter 8, however, there are reasons to be sceptical of this idea as a general principle for conceptual ethics.

4.3 Conceptual Tensions

Another set of theoretical vices comes into view once we focus on intra- and interconceptual tensions. When a concept gives rise to such tensions, a more authoritative alternative might be sought by re-engineering the concept to alleviate the tensions. If F and G are two concepts that are in tension with each other, for instance, and a re-engineered version of F , F' , would resolve the tension with G , a tidy-minded conception of conceptual authority would encourage one to conclude that F' is to that extent more authoritative than F .

Intra- and interconceptual tensions can take a variety of forms. As I propose to distinguish them, conceptual tensions can render concepts *unsatisfiable*, *inconsistent*, *incoherent*, *incongruent*, or *inimical*. This is also less well-trodden territory, and it will provide useful background to later chapters to map it out in more detail.

It may be said that, strictly speaking, concepts can never directly conflict or be in tension with each other—any talk to that effect must really be a shorthand for conflicts or tensions between the propositions or judgements that these concepts enable us to formulate, or else between the practical attitudes one takes towards courses of action ('*Vorrei e non vorrei*', Zerlina confesses in *Don Giovanni*—'I would like to and I would not like to').

But while it is right that tensions between concepts manifest themselves through conflicts between propositions, judgements, and attitudes, *which* propositions, judgements, and attitudes are properly accessible to one is a function of the concepts one uses, and the root of such conflicts *can* lie in an individual concept that systematically gives rise to them.

A helpful way of thinking about conceptual tensions is to conceive of the use of a concept as governed by certain proprieties of use—the norms or principles that mark the distinction between correct and incorrect applications of the concept by specifying *when*

the concept is applicable and *what follows* from its applicability.¹⁴ (I put it in terms of applicability, since that will serve us best for the thick normative concepts I focus on, but for certain types of concepts that I do not discuss here, such as connectives, the relevant aspects of use are more naturally characterized in different terms, as they have more to do with knowing how to handle and evaluate constructions in which these connectives figure.)

This picture of concept use as governed by norms or principles might be resisted on the grounds that most people would be hard-pressed to list all the principles that supposedly govern their use of a given concept. But the picture is not the Platonic or intellectualist one on which our practices of concept use merely implement a catalogue of explicit and antecedently given principles. Rather, from an explanatory perspective, the practices precede the principles: the explanatorily basic case is that in which the principles or—to put it in less intellectualist-sounding terms—the proprieties of use are *implicit* in a custom or practice of concept use. They are a form of know how: a practical competence to tell what is a reason for or against what, in particular by distinguishing situations where a concept applies from situations where it does not, and what follows from its applicability from what does not follow. The very idea of explicit principles only makes sense against the background of implicit proprieties of use that can inform the interpretation and application of explicit principles, since their interpretation and application remain underdetermined otherwise, and invoking further explicit principles to render them determinate would engender a regress.

This pattern of argument is familiar from Lewis Carroll's 'What the Tortoise Said to Achilles' (1895), which aims to show that the inference rules governing how to move between explicit postulates within a given logical system cannot exhaustively be expressed by explicit postulates *within that system*.¹⁵ Wittgenstein's rule-following considerations then reiterate the same pattern of argument one level deeper, suggesting that proprieties of inference cannot exhaustively be expressed in the form of explicit rules to begin with, because any given rule leaves its own interpretation and application

¹⁴ For a particularly illuminating way of setting out proprieties of use, see Michael Williams's (2013) notation.

¹⁵ Though see Besson (2018, forthcoming) for a detailed critical discussion of Carroll's argument and the contentious assumptions it makes.

underdetermined, and attempts to remedy this through the introduction of further explicit rules engenders a regress. Consequently, explicit, rule-based proprieties of use have to be grounded in implicit, practice-based proprieties of use.¹⁶

As mentioned in Chapter 1, moreover, we need not assume that people are always *consciously guided* by explicitly represented norms as they apply concepts. We may, most of the time, merely be actualizing our dispositions to conceptualize the world in certain terms and reason along certain lines, without conscious guidance by norms; and yet this actualization of dispositions is nonetheless *governed* by norms insofar as it is *liable to assessment and sanction* by fellow concept-users, who may enforce the norms by reproaching us when we apply a concept incorrectly.¹⁷ Like other norms, the proprieties governing concept use typically become manifest in their transgression. The situation of improper use is the characteristic situation in which others are prompted to make implicit norms explicit, thereby bringing to consciousness what may otherwise be unthinkingly followed.

But proprieties of concept use are also made explicit when a concept is taught to a novice. Modal vocabulary plays a crucial role in this, as it does in correcting improper concept use. Expressing and conveying norms has been thought to be one of the main functions of modal vocabulary.¹⁸ This comes out in how quickly we reach for modal language in teaching someone the key concepts of a game, for example: ‘White *must* move first.’ ‘The King *may* move backwards.’ ‘It is *impossible* to move backwards with a pawn’.

Analogously, mastering a concept centrally involves coming to understand what inferential moves one can properly make with it. There need not be an entire catalogue of norms looming under each concept that exactly and exhaustively determines its correct use across all imaginable situations. Taking it to be necessary would, as the Wittgensteinian analogy has it, be like supposing that, whenever children play a game

¹⁶ See Wittgenstein (2009), especially the strand among his many different considerations on rule-following that leads up to §201. See Brandom (1994, 22) for a concise exposition of the issue, and Brandom (1994, 18–46; 2019a) for a detailed defence of the explanatory priority of implicit over explicit norms; see also Kripke (1982), the essays in Miller and Wright (2002), as well as Kusch (2006) for discussions of this theme.

¹⁷ See Hlobil (2015) for a defence of this focus on assessability rather than guidance.

¹⁸ See Brandom (1994, 2015b) and Thomasson (2020a, ch. 2).

with a ball, they must be playing according to exact rules that strictly regulate every aspect of the game.¹⁹

All we require for our purposes is the idea that the use even of a single concept is liable to assessment according to *some* proprieties of use. And a plurality of such proprieties, even if implicit and non-exhaustive, already makes it possible for these proprieties to *conflict* by yielding incompatible instructions.

There are several different ways in which instructions can prove incompatible, both within and between individual concepts. Let us focus first on concepts harbouring *intraconceptual* tensions, i.e. tensions within a single concept.

Within the set of concepts harbouring intraconceptual tensions, we can then distinguish between, on the one hand, concepts that place incompatible demands on the world they are applied to, and, on the other hand, concepts that place incompatible demands on the concept-users that apply them.²⁰ We can mark this distinction by calling concepts that place incompatible demands on the world *unsatisfiable*, and concepts that place incompatible demands on concept-users *inconsistent*.²¹ Consider unsatisfiability first:

Unsatisfiability:

A concept is unsatisfiable iff nothing satisfies its conditions of application, because it places incompatible demands on the world in which it is deployed.

This can in turn either be due to the fact that its norms of application are incompatible a priori, or to the fact that, given what the world we inhabit happens to be like, nothing in fact ever meets these conditions. We can therefore distinguish two kinds of unsatisfiability:²²

¹⁹ See Wittgenstein (1958, 25).

²⁰ The distinction is flagged by Yablo (1993, 372).

²¹ This notably accords with the use of that terminology in Chihara (1979, 593) and Scharp (2013, 39), for example.

²² For a similarly bifurcated account of the possibility of ‘contradictory concepts’, see Priest (2014), who also defends the distinction between conceptual and worldly factors against Quinean qualms in Priest (2016).

A Priori Unsatisfiability:

A concept is unsatisfiable a priori iff its proprieties of use are such that nothing could possibly satisfy its conditions of application.

A Posteriori Unsatisfiability:

A concept is unsatisfiable a posteriori iff its proprieties of use are such that, in the world in which the concept is used, nothing ever satisfies its conditions of application.

An example of a priori unsatisfiability would be the concept *squircle*, whose use we might take to be governed notably by the principle that if x is a square and x is a circle, then x is a squircle.²³ Since nothing could possibly be both a square and a circle, the concept is unsatisfiable in principle. Further examples include *totalitarian democracy* and other concepts whose linguistic expression generates oxymoronic phrases.

An example of a concept that is unsatisfiable a posteriori is the concept *perpetuum mobile*, understood along the following lines: if x is a human-scale machine that can work infinitely without an energy source, x is a *perpetuum mobile*. Such a machine is not inconceivable in principle—many people have tried to build one—and some particles may even display repetitive perpetual motion on a microscopic scale; but given the laws of thermodynamics in the world we live in, no human-scale machine can satisfy the concept.

Unsatisfiable concepts contrast with *inconsistent* concepts, which place incompatible demands on the concept-user. They might give rise to paradoxes or contradictions, for example, by dictating that a concept both *applies* and *does not apply* in one and the same situation.²⁴ This engenders a tension within a single concept, which we may call an *intraconceptual inconsistency*, or simply *inconsistency* for short:

Inconsistency:

A concept is inconsistent iff its proprieties of use place incompatible demands on the concept-user.

²³ I take the example from Scharp (2013, 39).

²⁴ The possibility of inconsistent concepts is acknowledged and discussed in Bennett (2008, 57), Scharp (2013, 35–56), Eklund (2019), Pinder (2019), and Greenough (2020).

Consider Kevin Scharp's example of an inconsistent concept, the concept *rable*.²⁵ He characterizes its proper use as being governed by the following two principles: (i) if x is a table, the concept *rable* applies to x ; (ii) if x is a red object, the concept *rable* does not apply to x . We understand these principles well enough to deploy the concept *rable* without difficulty in most cases. But perplexity befalls us when we encounter a red table. Since it is a table, the concept applies to it; but since it is a red object, the concept at the same time does not apply to it. Therefore, the concept simultaneously applies and does not apply to the red table, which is inconsistent.

There are also less artificial examples of inconsistent concepts. As Graham Priest shows in his study of contradictions, the law offers numerous illustrations of inconsistent concepts.²⁶ A quasi-historical, simplified example obeying the same pattern as Scharp's concept *rable* is the concept *enfranchised* as defined as follows: (i) if x is a property-holder, the concept *enfranchised* applies to x ; (ii) if x is a woman, the concept *enfranchised* does not apply to x . As Priest notes, no blatant inconsistencies arise as long as no woman holds property in the jurisdiction in which this concept is in use. But the moment a woman becomes a property-holder, the concept suggests both that she is and that she is not enfranchised.²⁷

On Scharp's account, however, we need not even look beyond philosophy to find inconsistent concepts: philosophy itself revolves largely around inconsistent concepts. On his view, 'knowledge, nature, meaning, virtue, explanation, essence, causation, validity, rationality, freedom, necessity, person, beauty, belief, goodness, space, time, and justice' (2020, 397) have all turned out to be inconsistent concepts. He has argued in detail for the claim that the concept *truth* is inconsistent, and suggested that this is why the concept gives rise to paradoxes such as the Liar's paradox, Curry's paradox, or Yablo's paradox, and to contradictory conjunctions of the form ' p and not- p '.²⁸

However, inconsistencies should not be *identified* with the contradictory conjunctions they engender. When using an inconsistent concept, as Stephen Yablo has emphasized,

²⁵ See Scharp (2013, 36)

²⁶ See Priest (2006, 182–204).

²⁷ See Priest (2006, 184–5; 2014, 15).

²⁸ See Scharp (2013, 2020). On his account, these are three variants of the same paradox.

we may find ourselves flip-flopping back and forth between thinking that p and thinking that not- p without necessarily drawing the conclusion: ' p and not- p '.²⁹ We are then 'continuously being driven from one decision to the contrary one' (1981, §686), as Wittgenstein puts it. Inconsistencies of this sort are therefore best defined as residing not in the contradictory conclusion itself, but in the proprieties of use that lead to it. For any inconsistent concept, we can ask which *subsets* of proprieties or principles governing its use are consistent or inconsistent, and the inconsistency itself can then be located more precisely by identifying the smallest possible subset that still generates the contradictory conjunction.

Incompatibilities might arise at both ends of a concept, which is to say not only between the proprieties of use governing under what conditions the concept is properly applied, but also between the proprieties of use governing what properly follows from its application. The applicability of one concept might entrain two inconsistent obligations, for example. Or, in a Kafkaesque bureaucratic set-up or a catch-22 situation, a concept might be clearly and unequivocally applicable, yet give practical instructions that turn out not to be jointly satisfiable, because they mutually exclude or mutually presuppose each other. More generally, any concept that picked out a set of objects and then enjoined one to draw contradictory conclusions from the concept's applicability would be inconsistent in virtue of its consequences of application.

Such inconsistencies arising downstream of the concept's applicability remain inconspicuous as long as we think of concepts on the model of functions from worlds (or world/time pairs) to sets of things picked out, since this model tends to channel attention to the applicational face of the concept and away from its consequential face. But once we think of concepts as two-faced, having both conditions and consequences of application, we can distinguish two kinds of inconsistency:

Inconsistency in Conditions of Application:

A concept has inconsistent conditions of application iff the proprieties of use governing its application place incompatible demands on the concept-user.

²⁹ See Yablo (1993, 371).

Inconsistency in Consequences of Application:

A concept has inconsistent consequences iff the proprieties of use governing the inferential consequences of its applicability place incompatible demands on the concept-user.

These definitions of unsatisfiability and inconsistency capture four different ways in which a single concept can by itself already give rise to tensions by yielding incompatible instructions. They also bring out that the relevant instructions are sometimes world-facing and sometimes user-facing. Concept-users may have no difficulty in consistently employing a concept whose demands the world cannot satisfy, just as the world may have no difficulty in satisfying the demands of a concept that concept-users find it impossible to employ consistently—concepts can be unsatisfiable without being inconsistent, or inconsistent without being unsatisfiable.

Understanding how concepts can place incompatible demands on us downstream as well as upstream of their application is also key to seeing how different concepts can conflict. Two concepts may each yield instructions that are consistent in themselves, but that conflict when taken together. This gives rise to *interconceptual* tensions—conflict *between* rather than *within* concepts.

To a first approximation, we might say that two concepts conflict insofar as the instructions given by one concept conflict with the instructions given by another concept. But note that ‘conflict’ has to mean something stronger than mere *incompatibility* between two concepts—the fact that the applicability of one concept to x is incompatible with the applicability of at least *some* other concepts to x (if it is a *bird*, it cannot also be a *fish*) is not an optional feature of our conceptual apparatus that we might intelligibly seek to eliminate, but an utterly basic condition on the very contentfulness of thought. A concept whose applicability did not rule out—and, in that sense, ‘conflict’ with—the applicability of any other concept would be empty, not only in extension, but of content; just as a wheel that turns though nothing turns with it is not really part of the mechanism, a concept that applies without excluding the applicability of any other concept is not really part of the conceptual apparatus.³⁰ Even though the concept *bird* is incompatible with the concept *fish*, in that the applicability of the concept

³⁰ The wheel analogy hails from Wittgenstein (2009, §271).

bird rules out the applicability of the concept *fish*, the two concepts do not conflict as long as they do not claim to be applicable to the same object at the same time.

Two concepts do conflict, however, when the applicability of the one rules out the applicability of the other *and* their application conditions are such that they apply to the same object: if there were a concept that was at least partly co-extensional with the concept *fish*, but whose applicability excluded the applicability of the concept *fish*, it would conflict with the concept *fish*. Let us label that form of conflict—which combines co-extensionality with inferential incompatibility—*incoherence*:

Incoherence:

Two concepts are incoherent iff (a) the proprieties of use governing their correct application are such that the two concepts apply to some of the same objects *and* (b) the proprieties of use governing the concepts' inferential consequences are such that the applicability of one concept rules out the applicability of the other.

Besides incoherence, however, there is another form of interconceptual conflict that arises not at the level of *inferential* compatibility, but at the level of *extensional* compatibility: two concepts can interfere with each other's *instantiation*. They are not so much incoherent as *incongruent*, like two shapes that will not fit together into one box unless a chunk is taken out of at least one of the two shapes, thereby creating a remainder.

As with inferential incompatibilities between concepts, there is an utterly basic form of incongruence in any conceptual apparatus that is not an interesting feature of some concepts in contrast to others, but rather a condition on concepts having *distinct* extensions at all: the concepts *vanilla ice cream* and *chocolate ice cream* might also be said to be incongruent in the minimal sense that filling the universe with vanilla ice cream would leave no room to fill it with chocolate ice cream—but that is not a noteworthy tension between two concepts so much as a trivial consequence of the fact that they pick out different things.

Where incongruence becomes an interesting feature of some concepts in contrast to others is in the realm of concepts that are *action-guiding* in addition to (or, at the limit, instead of) being world-guided—in other words, concepts whose engaged use typically gives people *reasons for action*. In virtue of encouraging or prescribing certain courses of

action, these concepts give different practical instructions that can prove impossible to realize in concert. As a result, inhabitants of a conceptual architecture shaped by concepts that are incongruent in this sense will find that they face hard choices, because they cannot possibly live up to the demands these concepts make on them without leaving an unrealized remainder.

Paradigmatically, this type of conflict arises from the combination of different concepts of goods, values, virtues, or ideals—concepts such as *duty*, *honour*, *friendship*, *loyalty*, *truthfulness*, *magnanimity*, *genius*, *solidarity*, *generosity*, *equality*, *liberty*, *justice*, *modesty*, *asceticism*, or *humility*, to name but a few. These concepts clearly do not merely pick out something, but enjoin their engaged users to realize or instantiate something.

In some cases, the reasons for action these concepts introduce into the engaged user's deliberation will directly refer to what the concept is a concept *of*: the reasons one becomes sensitive to in coming to live by the concept of justice, for example, favour acting *in the name of justice*, or doing something *because it is just*; similarly, someone who lives by the concept of duty will characteristically do something *because it is her duty*.

But concepts may also exert their action-guiding influence indirectly, by rendering the person who lives by them sensitive, in her own practical deliberation, to aspects or considerations that make no mention of the concept with which an onlooker might describe the good or value realized by the ensuing action: someone who lives by the concept of generosity, for example, can act *out of generosity*, but doing so must precisely not involve being motivated by the consideration that *it is generous*; that would be to do something else—*posing* as someone who lives by the concept of generosity, perhaps, or else engaging in the sort of moral self-indulgence that consists in being motivated primarily by one's image of oneself as having certain moral virtues and dispositions.³¹ Another example is the concept of modesty: those who embrace the concept and live by it will do things or refrain from doing things *out of modesty*, but to do something *because it is modest* is to be motivated by a reflexive concern with one's own dispositions that borders on the hypocritical: it is a second-order substitute for a genuinely modest disposition, and one that is, if anything, immodest.

Whether they exert their action-guiding influence directly or indirectly, however, such

³¹ This notion of moral self-indulgence is introduced by Williams in the context of a discussion of utilitarianism (Williams 1981g).

concepts can conflict in the sense that the realization or instantiation of one concept comes at the expense of the realization or instantiation of the other. If, for example, one lives by two concepts of genuinely distinct and incommensurable intrinsic goods; and if these goods cannot be fully realized in concert; then the two concepts conflict in the sense that they will not fit together into one life, or one society, without remainder; and that remainder will represent not just an unrealized potential, but a *loss*: the real cost of realizing another good. One would not experience that loss if one lacked the concept of the good in terms of which the loss was incurred, or if that concept were different in the way required to render the good fully co-realizable with the other good. (These points will prove crucial in Chapters 5 and 9.)

I shall label this form of interconceptual conflict *incongruence*:

Incongruence:

Two concepts are incongruent if and to the extent that the realization or instantiation of one concept comes at the expense of the realization or instantiation of the other concept.

The phrase ‘if and to the extent that’ registers the fact that incongruence comes in degrees. Up to a point, one may be able to jointly realize or instantiate two action-guiding concepts without trade-off, and it is only under certain circumstances, or if the concepts are taken beyond a certain point, that the incongruence between them becomes manifest. To say that two concepts are incongruent is thus not to say that they always and everywhere conflict with each other, but that they are *set up* or *poised* to conflict under certain circumstances or when taken far enough.

We need to distinguish, however, between incongruities that are primarily the product of unpropitious circumstances and incongruities that are primarily the product of the concepts we use. It may be that two concepts are merely *accidentally* incongruent, because the world happens to put contingent obstacles in the way of their co-realization; or, more interestingly, it may be that two concepts would be incongruent even under ideal circumstances, because their incongruence cannot be remedied simply by augmenting the space and resources available to realize them, but is inherent in the concepts themselves, so that we struggle to so much as *conceive* of circumstances in which those

concepts would be always and everywhere realizable without loss.³² There is something about the respective logic of the concepts themselves that systematically tends to give rise to a tension between them. The concepts are, in that sense, *non-accidentally* incongruent, and the only way to eliminate all occasions for that incongruence to manifest itself is to change not the world, but our concepts.

Accordingly, non-accidental incongruities mark a type of interconceptual tension that is not simply a generic and ineliminable feature of all concepts, nor merely a product of unpropitious circumstances, but specifically a product of combining certain concepts in one's conceptual apparatus. And this type of tension is of philosophical interest, notably, because it is key to understanding *conflicts of values*, in the broad sense of 'values' that encompasses obligations, virtues, norms, aspirations, and ideals. In particular, this type of tension is key to understanding the competition over which values should be realized, the hard and occasionally tragic character of the choices involved, and the genuine losses and grounds for regret and complaint engendered by the frustration of one value for the sake of another.

One well-known example of such an incongruence is that between the engaged use of the concepts *truthfulness* and *happiness*, which Nietzsche registered as follows: 'there is no pre-established harmony between the furthering of truth and the well-being of humanity' (1986, §517)—or, as Lord Byron's tortured hero Manfred put it, 'The Tree of Knowledge is not that of Life' (2015, I, i, 12).

Another well-known example is the conflict between *liberty* and *equality*. The greater the liberty that individuals enjoy in a society, the harder it tends to become to secure equality between them. Conversely, increasing and enforcing equality tends to entail a cost in liberty. Modest gains in both liberty and equality may be achievable without trade-off, but when either value is taken far enough, there comes a point at which one has to be sacrificed to the other.

The same incongruence arises with a group of value concepts that has achieved renewed prominence in philosophy over the last decades, namely *virtue concepts*—concepts of ethically valuable dispositions or character traits whose possession renders the ethical significance of certain facts or states of affairs salient to the possessor. To be

³² For a rich discussion of this theme drawing on the work of Isaiah Berlin, Stuart Hampshire, and Bernard Williams, see Hall (2020).

an engaged user of the concept *honesty*, for example, is to value the disposition to honesty in oneself and others, to condemn and shy away from lying, and to try to live up to the ideal of honesty one embraces. But virtue concepts will conflict insofar as not all admirable dispositions or character traits can be equally embodied at the same time, or even by the same person (or the same institution, for institutions can have virtues and vices, too).³³ One may aspire to be both honest and kind, but find that, sometimes, one virtue can only express or manifest itself at the expense of the other. It may also be difficult to combine certain character traits *tout court*—for example, it is notoriously difficult for one person to cultivate both spontaneity and self-control, and it is well-nigh impossible to lead a life as both an ascetic and a hedonist.

That there were such incongruities between virtues is precisely what the ancient thesis of the unity of the virtues disputed (thereby providing a virtue-ethical example of the assumption of harmonious unity between concepts that we encountered in the previous chapter). Those who have since cast doubt on that thesis have insisted that virtues cannot always be combined—the demands that virtue concepts place on us conflict because the realization of one virtue interferes with the realization of another. For a social worker in an antipoverty agency, for example, the demands of efficiency and fairness will frequently conflict with the demands of compassion and generosity.³⁴ Though that incongruence will be aggravated by a lack of resources, it is not merely a reflection of such a lack. The administrative virtues of efficiency and fairness, on the one hand, and the Christian virtues of compassion and generosity, on the other, themselves pull in systematically divergent directions, both in terms of what they enjoin one to do and in terms of the spirit in which they enjoin one to do it. Where the former virtues demand professional detachment, restraint, and being stingy with one's time, the latter virtues demand empathy, caregiving, and going out of one's way to help. As one social worker sums up the problem: 'we're doing such a number game, and then we try to be a people person. You can't do both' (Zacka 2017, 203).

Finally, there are cases in which concepts are not just incongruent, but *inimical*, because one concept itself discourages or condemns the realization of the other. The

³³ For an elaboration of the claim that institutions themselves can have vices and virtues, see Fricker (2010a, 2020b).

³⁴ See Zacka (2017, 100).

concepts do not just interfere with each other's instantiation; rather, having one value concept constitutively involves *disvaluing* what having the other value concept constitutively involves *valuing*.³⁵ The ancient Greek value of *arete*, which enjoins one to achieve excellence in all things through self-cultivation, is not just incongruent with the Christian values of humility, selflessness, and abnegation of worldly glory, but actively discouraged and disparaged by them—just as these Christian values came to be condemned in turn by the Renaissance notion of *virtú*, which rehabilitated the pursuit of worldly glory and the prowess displayed by the *virtuoso*.³⁶ Nietzsche spoke advisedly in this connection of *reversals* of values, in which the polarity of valuations was inverted to produce not just *different* values, but *opposite* values.³⁷ Nietzsche would not have objected with such urgency to the Christian virtues had they not been inherently inimical to the realization of the more life-affirming pagan and Renaissance virtues he endorsed—and before him, Hume voiced much the same complaint about '[c]elibacy, fasting, penance, mortification, self-denial, humility, silence, solitude, and the whole train of monkish virtues' (1998, 9.3).

Concepts can thus exhibit of a variety of features that appear as theoretical vices to the tidy-minded: they can be unclear, ill-demarcated, referentially or inferentially indeterminate, vague, porous, gerrymandered, or unfruitful; and they can be tension-ridden in various ways, notably by being unsatisfiable, inconsistent, incoherent, incongruent, or even inimical. With a clear sense of the kinds of theoretical vices concepts can display, we are now in a position to consider the project of eliminating these theoretical vices from our conceptual apparatus: the tidy-minded pursuit of conceptual authority through theoretical virtue.

4.4 Authority through Theoretical Virtue

Tidy-mindedness, when applied to our conceptual apparatus, aims at a tidy mind: a mind whose thoughts are cast in theoretically virtuous terms—terms that are clear, precisely demarcated, fully determinate, fruitful, satisfiable, consistent, coherent, and congruous.

³⁵ See also Nagel (2001, 107), who calls these cases of 'true opposition' between values.

³⁶ On the Renaissance notion of *virtú*, see Skinner (2002, 2017) and Owen (2018).

³⁷ The phrase he uses in German is '*Umwertung der Werte*'. See Skinner (1997), Owen (2018), and Queloz (2021a) for discussions of such value reversals.

This can seem like a mere aesthetic preference, or even an ‘obsession’ (Geuss 2020, xviii), but it is more illuminatingly understood as a non-foundationalist expression of the pursuit of authoritative concepts, one that treats theoretical virtues as the answer to the authority question.

But where and to what extent is it appropriate to see conceptual authority as deriving from theoretical virtue? And *why is it* that greater precision, determinacy, or consistency should give concepts more authority?

Where the realization of theoretical virtues is the dominant concern anyway—in logic and formal semantics, say—questioning the authority of theoretical virtues seems moot, because the aim of engineering more theoretically virtuous concepts aligns with the aim with which the existing concepts were being deployed anyway. And it is primarily such contexts that Carnap has in mind when he first proposes, in 1947, that philosophers should engage in what he calls *explication*: the task of taking the less theoretically virtuous concepts of prescientific thought and engineering them to be more theoretically virtuous (1947, 7–8). For Carnap, a ‘concept must fulfil the following requirements in order to be an adequate explicatum for a given explicandum: (1) similarity to the explicandum; (2) exactness; (3) fruitfulness; (4) simplicity’ (1950, 5). But in spelling out the significance of these four desiderata, he makes clear that they are to be realized ‘so as to introduce the explicatum into a well-connected system of scientific concepts’ (1950, 7). His examples of successful explication are also confined to modern logic and scientific theorizing.³⁸ This is significant, for as long as conceptual engineering efforts aiming at more theoretically virtuous concepts are confined to enterprises that themselves aim to realize theoretical virtues in the form of systematic theories, the authority question does not arise: the concepts can straightforwardly draw their authority from the alignment between the spirit of the engineering project and the spirit of the enterprise whose concepts are being engineered.

But the authority question resurfaces once the aspiration to engineer for theoretical virtues is *generalized* beyond the confines of logic and formal semantics. The trajectory of Kevin Scharp’s work offers an illustration of this. In *Replacing Truth* (2013), Scharp proposes to replace the concept of truth, which he thinks is inconsistent and generates

³⁸ For a detailed reconstruction of Carnap’s conception of explication and the paradigms he is inspired by, see Carus (2007), Reck (2012, forthcoming) and Dutilh Novaes and Reck (2017).

various paradoxes, with two new concepts of truth; but he is clear that the substitution is to be confined to contexts in which consistency and the avoidance of paradoxes are the dominant concerns.

In his more recent methodological writings, however, he drops this qualification, conveying the impression that theoretical virtues can be a *general* answer to the authority question. Not just logic, but philosophy more broadly ‘is the study of what have turned out to be inconsistent concepts’ (2020, 398), Scharp maintains, because ‘truth, knowledge, value, virtue, freedom, justice, etc.’ have turned out to be ‘organized and distinguished by principles that are themselves inconsistent with one another’ (2020, 414). These inconsistencies generate various problems that philosophers get entangled in, and that conceptual engineering promises to resolve. But even if the project of engineering concepts is initially motivated not by the pursuit of consistency for its own sake, but by the desire to overcome the problems and paradoxes generated by our concepts, the engineering effort itself is guided by and aims for theoretical virtue. Across the entire range of our concepts, Scharp suggests, philosophy’s guiding ideal should be ‘a consistent conceptual scheme. No paradoxes. No puzzles. Just clarity’ (2020, 415).

On this view, which presents us with a paradigmatic embodiment of the tidy-minded approach, philosophical theories are to be cast as *measurement systems*, so that our messy everyday judgements involving some concept can be transposed into a more rigorous, precisely defined, and consistent language. Scharp calls this view *metrological naturalism*:³⁹

Metrological naturalism has as a methodological principle that philosophers should use measurement theory as a guide or model in philosophical theorizing. ... [W]e know pretty well how to do this for things like length and weight. Trying to figure out how to construct a measurement system for something like truth or justice is a lot more complicated, but this isn’t just an analogy. (2020, 402)

He maintains that engineering should always be conducted in a metrological spirit, because ‘engineering without metrological naturalism is blind’ (2020, 399): the virtues of a good measurement system are what provides engineers with a guiding sense of what

³⁹ Examples include Davidson’s (1990) measurement system for belief, desire, and meaning. See also Matthews (2010) and Weaver and Scharp (2019).

concepts to aim for.

But how plausible is it, outside logic or formal semantics, that an engineered concept will be authoritative because it exhibits the virtues of a good measurement system? Consider the concept *person*. In its colloquial use, the concept is vague, and its connection to other concepts not very systematic. It indicates a variety of characteristics—self-consciousness, agency, title to respect—that come in degrees, and, as debates over abortion show, its ethical implications and relations to other categories such as *sentient being* or *human being* are indeterminate and contested. Some, like Michael Tooley, have therefore undertaken to engineer it into a precise sortal notion that sharply delineates a basis for a more systematic way of thinking about issues surrounding personhood.⁴⁰ Yet the implications of this precise sortal notion notoriously go drastically beyond anything within the reach of the non-engineered concept. Granted, making a notable difference to the resulting judgements is part of the point of engineering the concept. Faced with a stark divergence between pre-engineering and post-engineering judgements, however, the question of the engineered concept's authority becomes acute: why should we act on the judgments formed using the engineered concept? Because it is more precise and permits a tidier way of thinking? But say the engineered concept licenses infanticide in situations in which our non-engineered concept suggests that infanticide is abhorrent. Why should we care about the added tidiness when its price is to do something which, from the perspective of ingrained ethical experience, appears deeply revolting? If that is the price of tidiness, one may well think, then too bad for tidiness.

What comes out here is that displaying theoretical virtues is not the only thing we need our concepts to do.⁴¹ And once this much is granted, it becomes an open question whether vague, indeterminate, open-textured, or tension-ridden concepts might not sometimes serve us best. Perhaps theoretical vices have other virtues. The tidiest

⁴⁰ See, e.g., Tooley (1972, 1983) and many of the positions discussed in Merrill (1998), and see Williams (2011, 127) for a critique of Tooley on similar grounds. For an exposition of arguments over the definition of 'person' against the background of constitutional disputes over abortion, see Schiappa (2003, 89–108).

⁴¹ Herman Cappelen registers this possibility in describing W. V. O. Quine's views: 'there is no reason why there should be a fixed set of theoretical virtues that are used to measure improvement. In certain contexts, non-theoretical virtues/advantages could make a big difference' (Cappelen 2020, 137–138).

organization is not always the most functional, and the sharpest tools are not always the most suitable. If your concern is to cut bread and you ask for a bread knife, you will hardly thank me if I give you a razor blade because it is sharper.⁴²

In contexts in which we are not obviously, or not primarily, concerned to realize theoretical virtues, it is thus not clear that concepts engineered to be free of theoretical vices will carry more authority than the concepts they are meant to replace. Concepts do not possess more authority *simply* by dint of their theoretical virtues. What authority they have must come from the way they tie in with what is important to us. Sometimes, that is indeed tidiness, measurability, or mathematical rigour. But sometimes it is not. Concepts ‘*supply lacks*’, in Quine’s pithy phrase, and the character of what we lack is as variable as the concerns that make us lack it.⁴³

Thus, conceptions of conceptual authority that isolate concepts from the practical contexts in which they are put to work and concentrate on the inherent defects of concepts—or what appear as defects when measured against some ideal of theoretical virtue—embody a strategy that may satisfactorily answer the authority question in special cases, but that cannot hope to do so more widely. By focusing on the theoretical vices of concepts, we risk overlooking their other virtues, and might end up rendering the concepts less helpful than they were before we ‘ameliorated’ them. Striving invariably to realize theoretical virtues in our concepts threatens to be counterproductive when we are better served by concepts whose virtues are nothing like those of a tidy theory. What is more, it leaves unanswered the question of why we should care about those theoretical virtues in a given context.

4.5 Inheriting Authority from Theories

If the authority question persists after a concept’s theoretical virtues are invoked, it may

⁴² The example is Wittgenstein’s (2000, MS 120, 142v).

⁴³ When Sally Haslanger engineers revised versions of social and political concepts such as *gender*, *man*, *woman*, or *race*, for example, she does not try to make the case for her re-engineered concepts in terms of their theoretical virtues. Instead, she highlights the way in which these concepts promise to serve antecedent social and political concerns—concerns that precede, direct, and give point to the conceptual engineering effort itself. See Haslanger (2012). For a retrospective assessment of her conceptual engineer effort which also briefly raises the question of the authority of such an effort, see Haslanger (2020a, 231–237).

be thought that this is because each concept is being myopically considered in isolation. Perhaps, to understand how conceptual authority can derive from theoretical virtue, one has to step back from individual concepts and look at the wider structures they form—that is, at the *theories* that are the paradigmatic bearers of theoretical virtues. This points to a conception of authority on which a concept is authoritative not by dint of the theoretical virtues it realizes all by itself, but by dint of being integrable into a wider theoretical structure whose own claim to authority rests on the degree to which *it* realizes theoretical virtues. In other words, a concept inherits its authority from the wider theory in which it is embedded.

The connection between concepts and theories is particularly explicit in the so-called ‘Theory-Theory’ of concepts, which is informed by Thomas Kuhn’s (1970) work on paradigms and theory change.⁴⁴ Some exponents of the Theory-Theory think of concepts as being themselves partial theories embodying explanations of the relations between the overall theories’ constituents,⁴⁵ while others hold that concepts are to be individuated in terms of theories.⁴⁶ But this is distinct from a tidy-minded approach to conceptual authority in that it conceives of the connection between concepts and theories primarily as a descriptive claim about how to understand the concepts we have already. This has implications for conceptual authority, but they are indirect: it informs our working picture of our conceptual apparatus, and thereby affects what we look for and which questions we are disposed to raise. The more we view our concepts through a theory-shaped lens, the more our picture of our concepts comes to resemble the holistic picture we encountered in the previous chapter, and the more it invites the same doubts as to why, outside the carefully regimented silos of scientific disciplines, we should expect our inherited concepts to form such a tidy structure.

Yet the tidy-minded approach to concept appraisal gives the connection between concepts and theories a more normative significance: it turns it into the idea that our thought *should* take the form of a theory—a tidy body of knowledge organized in a way

⁴⁴ The ‘Theory Theory’ label goes back to Morton (1980).

⁴⁵ See, e.g., Keil (1989, 281): ‘Most concepts are partial theories themselves in that they embody explanations of the relations between their constituents, of their origins, and of their relations to other clusters of features’.

⁴⁶ See, e.g., Carey (1985, 198): ‘Concepts must be identified by the roles they play in theories’.

that revealingly corresponds to the structure of its subject matter. This aspiration goes back to antiquity, and is most closely associated with scientific bodies of thought.⁴⁷ Systematicity has been thought to be the hallmark of ‘science’ in the broad sense of *Wissenschaft*.⁴⁸ The concept *phlogiston*, for example, derived what authority it had in the eyes of eighteenth-century chemists not from the virtues of the concept itself, but from the fact that it formed the linchpin of a theory of combustion which promised to offer a tidy account of what could be observed—until fatal problems for the theory were identified.⁴⁹

But the same tidy-mindedness can be found in the sphere of practical reasoning. In the *Protagoras*, Socrates advocates the pursuit of the complete systematization and unification of practical reasoning into a *techne*—a *science* or *measurement system*—that is geared towards a single, antecedently specifiable and measurable end.⁵⁰ The dialogue also articulates a rationale for this pursuit: turning practical reasoning into a *techne* will help shelter human beings from *tyche*, the unpredictable play of fortune, by giving them more control over whether their lives go well.

In line with this aspiration, moral philosophy has a long tradition of validating or invalidating concepts according to whether they can be harmoniously integrated into an ethical theory. What is nowadays called ‘virtue theory’ is in some respects a poor example of this, since what the ‘theory’ tends to suggest we need to live well are virtuous dispositions, not a theory about them, and having these virtuous dispositions does not obviously involve thinking in terms of a theoretical structure at all.⁵¹ But Kantian and utilitarian theories are paradigm examples of attempts to discriminate between more and less authoritative ethical concepts by tidying up our ethical thought until it forms a virtuous theoretical structure.

⁴⁷ See Williams (2006d, 27–28) for a discussion of the various forms that this ambition to systematize took in Greek philosophy. For historical overviews of the role of the ideal of systematicity in philosophy, see Ritschl (1906) and Rescher (1979, 2005).

⁴⁸ For a thorough vindication of the claim that systematicity is the hallmark of scientific knowledge, see Hoyningen-Huene (2013), which identifies nine different respects in which science is particularly systematic.

⁴⁹ See Blumenthal and Ladyman (2017).

⁵⁰ See *Protagoras*, especially 356d–e. For a fuller exegetical discussion, see Nussbaum (2001, ch. 4).

⁵¹ As emphasized by Williams (1995a, 551; 1996, 31; 1998, §1).

Of course, a theory imbuing individual concepts with authority is not necessarily non-foundationalist. It will invite a foundationalist interpretation insofar as it presents itself as resting on some universal and irrefragable foundation that it uses to overturn whatever conflicts with it. But when a theory presents itself as a systematization of our local concepts that is permitted to reflect the peculiarities of its local starting point, it invites a *non-foundationalist* interpretation. What is then supposed to ground the authority of concepts embedded within the theoretical structure are the theoretical virtues of an avowedly local structure of thought. This will then be a non-foundationalist conception of authority, because the authority in question remains authority *for us* in the bounded sense that contrasts ‘us’ with inhabitants of different social worlds who might reasonably object to the starting point of our theory construction.

Whether a given theory is foundationalist or non-foundationalist cannot always be read off the theory itself, but often only becomes fully determinate in conjunction with a certain interpretation of the theory. One and the same set of ideas can sometimes be interpreted either as grounded in timeless rational foundations or as embodying sociohistorically local standards. A case in point is the theoretical structure erected by John Rawls in *A Theory of Justice*. It invites one to take up the ‘Original Position’, in which a veil of ignorance masks all knowledge of contingent circumstances. To reflect on the human situation from the Original Position is ‘to see it *sub specie aeternitatis* … to regard the human situation not only from all social but also from all temporal points of view’ (1971, 587). This exercise is meant to yield identical and universally applicable results for any rational agent, and it is not least this foundationalist streak in Rawls’s earlier work that communitarian and realist critiques of political philosophy reacted against.⁵²

But Rawls’s later book, *Political Liberalism*, retains the theoretical structure of his earlier work largely unaltered while radically reinterpreting what it is a theory *of* and who it is a theory *for*: the Original Position is recast as a model for negotiating basic societal structures under constraints that are specifically those of modern liberal democracies. While the earlier Rawls sees his conception of justice as fairness as a solution to a timeless problem, justifiable in principle *to anyone*, the later Rawls sees it as a solution to the local

⁵² For communitarian critiques, see MacIntyre (1978, 1988, 2007), Sandel (1981, 1996), Taylor (1985, 1989), and Walzer (1983, 1987). For realist critiques, see Williams (1985, 1989, 2005f) and Geuss (2008) as well as Forrester (2019) and Hall (2020) for broader overviews.

problem of how to arrive at a conception of justice that is justifiable *to us today*, in societies facing unprecedented ethical diversity and difficulty in reaching substantive agreement.⁵³

Even when recast in non-foundationalist terms, however, theories such as Rawlsian contractualism, utilitarianism, or Kantianism retain one important commonality with foundationalism: they arrogate to themselves the authority to override the force of pre-theoretical concepts. On a kaleidoscopic picture of our conceptual apparatus, this is an authority they are bound to have to exercise, because building a tidy theory out of a messy and disparate conceptual inheritance forces one to overturn the authority of *some* concepts in order to establish the authority of *any* concepts. Moving from a collage of cross-cutting conceptual claims to a tidy and tension-free edifice requires prioritizing or ranking the claims our concepts make on us, using some of them to overturn others, and discarding or de-authorizing all the concepts and correlative reasons that cannot be integrated into the theory. The resulting structure then allows us to ratify or authorize all the reasons that can be understood as applications or extrapolations of the reasons provided by the applicability of the concepts at the heart of the theory; but it also requires that we winnow out all the reasons that cannot be so understood.

These theories would not be as influential as they are if their master concepts' claims to authority did not possess some plausibility when considered in isolation. But these claims also have a sharp end: their exclusionary implication that *no other concept* should have authority over our lives except insofar as it can be authorized by and shown to be an application of the theory. Once one appreciates the full force of this radically revisionary implication, the authority question becomes acute. If theory-building involves paring down and thinning out our conceptual apparatus in the name of building a theoretically virtuous structure, this pits the authority of the theory against the combined authority of all the concepts that do not fit into it. And if the theory is seen to overturn the claims to authority of an extraordinarily wide and dense array of thick normative concepts that have force with us, even though they may not be foremost in our minds as we philosophize in thin and general terms, the question of the *theory's* authority really does become a live question.

⁵³ See Rawls (1993, xviii); I draw on Queloz and Cueni (2021, 7) here. See Testini (2020) for a fuller discussion.

Note that the question is what authority the virtues of the theory bestow on the theory within one's practical deliberation, which is significantly different from the question that T. M. Scanlon (1992) hears in the work of Bernard Williams (2011) and Michael Walzer (1983, 1987). On Scanlon's rendering, they reject the claim to exclusive epistemic authority in all matters moral that they see some moral theorists as making, namely the claim that giving advice and correcting people's moral beliefs is the preserve of moral experts marked out as such by their possession of a moral theory. Scanlon grants that this is a questionable basis for a claim to such exclusive authority, but argues that few moral theorists, apart from 'some utilitarians' (1992, 4), make such a claim to authority. The authority question that I am concerned with here, however, does not primarily aim to question the epistemic authority of the theorist over other people, but the normative authority, within one and the same person's practical deliberation, of a theory that aspires to crowd out or override all concepts and considerations that it cannot assimilate to itself.

Granted, recasting one's thought as a tidy theoretical structure ensures that it exhibits a variety of theoretical virtues, and one could regard these theoretical virtues as intrinsically valuable—many regard the majestic simplicity of a theory as a source of beauty and awe, for instance. But appeals to the authority of theoretical virtues alone invite the Lichtenbergian worry that the majestic simplicity of the theory might reflect no more than the rather less majestic simplicity of the theorist. Similarly, the seductive clarity, coherence, and economy of conspiracy theories should not be mistaken for genuine understanding, just as the standardized, quantifiable metrics of value deployed by administrative bureaucracies should not be mistaken for the richer set of values they are meant to render tractable.⁵⁴ The mere fact that some structure exhibits theoretical virtues is not by itself enough to ground its authority. More needs to be said about why the concepts, reasons, and judgements acting as tent poles for the theory authenticate it as an *authoritative way* of realizing theoretical virtues.

Even if, in addition to recognizing the intrinsic value of theoretical virtues, more *can* be said to show that the theory is in fact sustained by concepts that have force with us, the crucial question will still be why we should care more about these concepts and correlative reasons forming a theoretically virtuous structure than we care about all the

⁵⁴ On the dangers of seductively clear systems in connection with conspiracy theories and bureaucracies, see Nguyen (2021).

other concepts and reasons at the expense of which this is achieved. Why should it matter so much that, as I go about my daily business and engage in private, personal deliberation, my patterns of thought exemplify theoretical virtues? After all, as Williams remarks, 'I'm not living my life in order to exemplify a mathematical theory' (1999, 246).

The issue can also be put in terms of confidence. Theories such as Kantianism, utilitarianism, or contractualism require us to be supremely confident in the small handful of concepts, reasons, and judgements that sustain them—so confident, indeed, as to be willing to deploy them to overturn any concept, reason, or judgement that conflicts with them.⁵⁵ But is such a completely one-sided distribution of confidence really reasonable? Is it not *unreasonable*, in the thick sense of the term that invokes proportionality and good sense, to place all one's confidence in a small handful of ideas and take everything else to follow from that?⁵⁶ Recognizing the authority of the thought 'suffering is bad' while continuing to recognize the authority of other ethical concepts and reasons against which that thought can be balanced is one thing. It is quite another to recognize *nothing but* the authority of that thought and deny authority to any concept or reason that cannot be derived from it and shown to be an application of it. What renders the authority question so acute in connection with these tidy-minded theories is not so much what these theories fundamentally affirm as what they are willing to deny on that basis—what they are willing to sacrifice to tidiness.

Of course, one may be willing to make this sacrifice if one regards the demand for tidiness or systematicity as a ubiquitous and overriding demand of rationality itself. But many are not prepared to accept this equation of the authority of theory with the authority of rationality without further argument, objecting that it betrays, in Williams's phrase, an overly *rationalistic conception of rationality*.⁵⁷

Williams's suggestion is doubly topical in the present connection, since it encourages us to conclude that the question of the authority of theory, which was meant to answer the question of the authority of certain concepts, itself boils down, ultimately, to the question of the authority of a concept, namely a certain conception of rationality—in particular, one on which the tidy-minded recasting of thought into the form of a theory

⁵⁵ A particularly pronounced example of this is Kagan (1989).

⁵⁶ See Williams (1999, 245).

⁵⁷ See Williams (2011, 20, 112–13, 124).

is nothing less than a ubiquitous requirement on thinking rationally at all, because the demand to tidy up thought is inherent in the very idea of rationality. If concepts draw their authority from a theory, the theory in turn draws its authority from a very specific conception of rationality. It is by this slim lever that the vast structure of our entrenched moral sentiments and convictions is to be levered out.

What might give one reason to accept this specific conception of rationality? One reason might be that when our concepts conflict, the rational as opposed to arbitrary resolution of conflict requires some *general currency of reasons* (articulated in terms of a universal metric of utility or a special notion of moral obligation, for example) by which the claims that various concepts make on us can be compared and weighed against each other.

But, as numerous critics have pointed out, we are perfectly capable of rationally resolving conflicts by exercising our judgement in the particular case without relying on a general currency of comparison; that does not make the resolution arbitrary: we still have *reasons* to resolve conflicts one way rather than another.⁵⁸ To say that something is a matter of judgement is not to say that it is a matter of arbitrarily plumping one way or the other—unless we are only prepared to count something as a genuine reason if it can be shown to be derivable from a theory, and to do that would be to presuppose rather than to lend succour to a rationalistic conception of rationality. Absent a compelling argument for why we should accept the rationalistic conception of rationality, we have reason not to lose sight of the sense in which one can think rationally without thinking in terms of a theory.

In response, it is often said that to settle for the kaleidoscopic jumble of concepts we happen to find makes it all too easy to provide reasons for one's unreconstructed judgements, and thereby renders reflection overly conservative. It is *overly* conservative, in particular, because the conceptual apparatus we start out from is likely to be in various ways hidebound, cruelly superficial and prejudiced, ideologically distorted, and already

⁵⁸ For an especially trenchant elaboration of the idea that conflict-resolution requires no more than judgements in the particular case, see Dancy (2004). See Berlin and Williams (1994, 306–7) for a defence of the point that choices between incommensurable values are not unreasonable. A sophisticated account of choices between incommensurable values is provided by the work of Ruth Chang (2002, 2015, 2016); see also the essays by various authors in Chang (1997).

teeming with simplistic attempts at tidiness and ill-considered theories. Accordingly, the argument runs, we need well-considered theories to help us diagnose distortions and drive out ill-considered theories.⁵⁹

But while it is certainly true that theories quickly become radically revisionary, it is not true that reflection can *only* be radically revisionary when it takes the form of a theory. Even when it forswears the machinery of theory, critical reflection still has the resources to unmask prejudices. In fact, there is an important sense in which it has *more* resources when it does not have to be couched in the terms of a tidy theory, since it can then draw not just on the concepts that have a place in the theory, but also on all the concepts that the theory sacrificed to tidiness—a collection of critical resources which typically includes not only a rich battery of thick ethical concepts, but also cultural, psychological, historical, political, and socio-economic concepts that are often much better suited to identifying self-serving prejudices as such. The racist or sexist may offend against a formal principle of universalizability by not applying reasons equally, for example, but, as Williams has emphasized, that rarely captures what is chiefly amiss in concrete instances of discrimination.⁶⁰ The reasons that the racist or sexist gives for his discriminatory behaviour, which a theory-driven critique would focus on, are often mere rationalizations that are believed, if they are, only because they further selfish interests. Appeals to the canons of rationality might hope to bring out how the rationalization was subtly irrational, but they would not even begin to bring out how it was blatantly dishonest. To get at the psychological and social forces that can help us understand what is really going on (as opposed to what the ostensible justification *pretends* is going on), critical reflection needs to become more detailed and concrete rather than more abstract and principled—which is to say that it should move in a direction that is precisely the opposite of that which issues in the thin and general categories of a tidy theory.

Once it is acknowledged that theory-building is neither the only nor even necessarily the best way of obtaining critical leverage over ordinary thought, the costs of theory-building become more salient. A number of philosophers have felt that the tidiest theories leave us with too few concepts to make sense of our immediate experience: like

⁵⁹ See, e.g., Nussbaum (2000, 70).

⁶⁰ See Williams (2011, 130) and Cueni and Queloz (2021). This is one of several respects in which theories' 'rhetoric of radical rationality conceals how conservative they are' (Williams 1995m, 183).

filters laid over experience, they blot out all but a few features of a situation—whatever cannot be captured in terms of moral obligations, or in terms of its impact on the utilitarian calculus, is screened out.⁶¹ This may make the challenges we face in our practical deliberations simpler, more quantifiable, or more tractable. But it also flattens our perception by eliminating nuances and dimensions of value. Faced with a choice between a conceptual apparatus that is theoretically virtuous but impoverished and one that is untidy but richly expressive of a wide variety of human concerns and dimensions of value, is it so clear that we should go with the former?

A popular compromise between these two alternatives is to say that we do not necessarily need full-blown theories, but we still need to subject our inherited concepts to the progressive pull towards theory and should aim for a *reflective equilibrium* between the conservative weight of pre-theoretical reasons on the one hand and the demands of systematicity on the other.⁶² Instead of building a theory from the master concepts down, we start by treating the welter of our inherited concepts and reasons as *constraints* on the demand for a tidy system of thought; we then extract what organizing concepts and generalized principles we can from them, and inch, through the mutual adjustment of particular judgements to more general principles and general principles to particular judgements, towards a *tidier* structure of thought.

Since such a process is thought to be capable of instigating the revision of inherited concepts and the construction of novel concepts, the reflective equilibrium approach can be cast as an answer to the authority question, capable of differentiating more from less

⁶¹ See Williams (1996, 15; 2011, 130) and Chappell (2015). For discussions of Williams's critique in particular, see Smyth (2019) and the essays in Heuer and Lang (2012). Hämäläinen (2009) relocates Williams's critique of theory in contemporary analytic ethics, while Chappell (2009) offers a Williamsian critique of ethical theory.

⁶² Ethical and political theorists frequently present their preferred theory as marking a reflective equilibrium between the demands of systematization and our pre-reflective intuitions or commitments—see Rawls (1971, ch. 1, §4–9) and Daniels (1979, 1996), for example—and the notion is also invoked in other parts of philosophy: Elgin (1983, 1996, 2017) advocates its use in epistemology, while Lewis (1983c, x) and Keefe (2000, ch. 2, §1) even defend its applicability to all of philosophy. See also Brun (2014), Tersman (2018), Daniels (2020), Rechmitter (2022), and Beisbart and Brun (manuscript) for discussions of the method's career and elaboration.

authoritative concepts even in the absence of a full-blown theory.⁶³ This is certainly true to its origins. Nelson Goodman, who first introduced the method in his discussion of the validity of inference rules in *Fact, Fiction, and Forecast*, saw the method as answering, fundamentally, to the problem of which concepts we should use. Goodman frames the problem as a projection problem: which concepts advert to properties that are inductively projectable from a sample to a larger population? He compares the ‘task of formulating rules that define the difference between valid and invalid inductive inferences’ to the ‘task of defining any term with an established usage’ (1983, 66). In both cases, we should proceed by dual adjustment according to the following principle: ‘A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend’ (1983, 64, emphasis removed). As Georg Brun (2020) has shown, moreover, the method of reflective equilibrium is instructively linked to the method of ‘constructive definition’ that Goodman had developed in an earlier book, *The Structure of Appearance* (1977), a method which closely resembles Carnap’s method of explication. Goodman’s method of reflective equilibrium is thus offered as a non-foundationalist solution to the problem of which concepts to think with. What is more, it is well equipped to acknowledge that tidiness has a price, and that gains in tidiness need to be weighed against our confidence in our unreconstructed concepts and judgements.

However, that method still fundamentally adheres to the view of conceptual authority as deriving from theoretical virtue, and hence still invites the question of the authority of those theoretical virtues. As Brun acknowledges, ‘[v]irtues of theories are indeed the driving force’ behind the method of reflective equilibrium, ‘because they motivate the transition to a theoretically more suitable system of concepts in the first place’ (2020, 950). When we assess the authority of a partially systematized set of concepts, we are bound to do so by reference to our everyday judgements and our pre-theoretical sense of the life that this set of concepts is supposed to help us to lead. And in many cases, as the example of the tidied-up concept *person* and its questionable implications for infanticide illustrated, the reasons that guide and flow from our unreconstructed conceptual apparatus count for more than the gains in theoretical virtue. Why should the product of a procedure which, in the name of theoretical virtues, overturns judgements and

⁶³ See Brun (2020; 2022, 15–16) and Rechnitzer (2022, 29, 51).

concepts that have force with us be granted more authority over our lives than the messier outlook that better matches our initial distribution of confidence? The onus is on the side of the systematizers to show why the reasons yielded by more theoretically virtuous concepts should be granted more authority than the reasons we immediately have for thinking as we do.

Again, in the context of scientific thought, it is plausible that the authority of theoretical virtues *can* be grounded in the assumptions and aspirations driving scientific inquiry. To put a very complex issue rather crudely: the systematicity of scientific thought is taken to count in favour of its authority because that thought is taken to aim to reflect the systematicity of the world it describes. Within such an attempt to describe a single, systematic world—what the Stoics called the *systema mundi*—theoretical virtues can operate as criteria of authority because they are, however defeasibly, criteria of truth.⁶⁴

But this line of argument for the authority of theoretical virtues does not obviously generalize to the thick normative concepts—ethical, political, legal, cultural, and aesthetic—that sustain, and are sustained by, a social world. These are not obviously part of an attempt to describe a single, systematic world. There are many social worlds, and there is no particular reason to presume of any one of them that it is, or should be, unified and systematic enough to turn theoretical virtues such as consistency and coherence into reliable criteria of conceptual authority. As Thomas Nagel puts it, ‘truth in science, in mathematics, or in history has to fit together in a consistent system’, but ‘[o]ur evaluative beliefs are not part of the attempt to describe a single world’ (2001, 108–9).⁶⁵

Canvassing reasons for abandoning our untidy ‘folk’ concepts for more theoretically virtuous concepts, McPherson and Plunkett offer the consideration that the *class* of concepts displaying theoretical virtues boasts an excellent track record of utility in inquiry, and if adopting more theoretically virtuous concepts has proven useful for scientific inquiry, one might think that it could prove similarly useful for what they call

⁶⁴ See Rescher (2005) for a detailed case to that effect.

⁶⁵ See also Hämäläinen (2009, 548), who urges us ‘to remove from our picture of moral theory an assumption concerning the relationship between systematic theoretical articulation and action-guidance’, namely the assumption that moral theories are pictures of a moral reality that was systematic all along. If we abandon this assumption, we will be more receptive to the idea that theoretical virtues ‘do not necessarily mirror a proper orientation in the moral realm’, and that moral theories are better viewed as ‘a box of tools to be used … to elucidate different aspects of morality’, much like literary theories.

'normative inquiry':⁶⁶

But, as they also go on to acknowledge, the persuasiveness of this argument again depends on how much like scientific inquiry 'normative inquiry' is taken to be—on whether there is an antecedent systematicity for our normative concepts to reflect, or whether 'normative inquiry' is better thought of as a combination of, on the one hand, deductive and ampliative reflection on what the concepts we are confident in imply within the social world in which we deploy them, and, on the other hand, critical reflection on the reasons we have to adopt or abjure those concepts and their concomitant reasons. If the judgements we form in normative inquiry are not 'part of the attempt to describe a single world', in Nagel's phrase, it is not at all clear that the authority of the concepts articulating them should primarily, or mainly, be grounded in their exemplification of a catalogue of theoretical virtues. There will then be no reason to think, and many reasons to doubt, that this is what normative inquiry is primarily about.

In this chapter, I have examined and ultimately found inconclusive the different forms of tidy-minded non-foundationalism that look to theoretical virtues as a general answer to the authority question. In each case, we ended up with some version of the question of what authority these theoretical virtues themselves possess in connection with thick normative concepts, and the attempt to ground that authority in the constitutive requirements of rationality was found to hinge on a rationalistic conception of rationality whose authority is itself in question.

The conclusion this leads us to is not that we should never systematize, but rather that the merits of systematization will depend on the answers we can give to the further questions of *when* and *why* we need to systematize. The authority of concepts ultimately has to come not from how concepts relate to each other, but from how they relate *to us*. To supplement more theoretically virtuous concepts with the grounds on which they can claim more authority, and to discriminate contexts in which these grounds are given from contexts in which they are not, a more comprehensive approach is needed—one that appraises concepts not just on the basis of their inherent properties or the systematic relations between them, but on the basis of how they tie in with our concerns.

⁶⁶ See McPherson and Plunkett (2020, 281).

CHAPTER FIVE

Concepts and Concerns

5.1 The Dworkin–Williams Debate

In the fall of 1998, a year after the death of Isaiah Berlin, the New York Institute for the Humanities convened a two-day conference in Manhattan to examine his intellectual legacy. The conference attracted an unexpectedly large audience and was covered by *The New York Times*.¹ It was also the scene of a coruscating debate between Ronald Dworkin and Bernard Williams, itself the culmination of a dialogue going back to seminars they held together at Oxford in the late eighties (as part of a series informally known as ‘Star Wars’).² In their various writings over the years, but most emblematically in their exchange in New York, Dworkin and Williams came to personify two rival approaches to conceptual tensions. Once their elliptical and compressed remarks in New York are connected to the scattered writings in which they developed their points and engaged with each other’s work, a sustained debate between Dworkin and Williams emerges, one that is focused and protracted enough to merit being labelled ‘the Dworkin–Williams

¹ Such was the demand for transcripts afterwards that the organizers were persuaded, against their original plans, to publish the presented papers together with the subsequent discussions. Aileen Kelly, Steven Lukes, Mark Lilla, Robert Silvers, Ronald Dworkin, Bernard Williams, Thomas Nagel, Charles Taylor, Avishai Margalit, Richard Wollheim, and Michael Walzer all presented papers, while Frances Kamm, Richard Bernstein, Fritz Stern, Neil Rudenstine and others contributed to the discussions. See Rothstein (1998) as well as Lilla, Dworkin, and Silvers (2001) for an account of the conference and its reception.

² See Guest (2013, 17, 247n20).

debate'.³

This underappreciated Dworkin–Williams debate holds several valuable lessons for the needs-based approach to concept appraisal I want to develop. I reconstruct the debate here as a springboard to that approach, before returning to the debate in a later chapter to explore some of the more complex issues it raises.

What is fundamentally at issue in the Dworkin–Williams debate is how we ought to deal with a well-known conceptual tension: the conflict between the values of *liberty* and *equality*. As Isaiah Berlin describes the conflict, ‘total liberty for the wolves is death to the lambs, total liberty of the powerful, the gifted, is not compatible with the rights to a decent existence of the weak and the less gifted’ (2013d, 12–13). Conversely, the thorough enforcement of total equality carries severe costs in liberty, which has to be curtailed to redress or prevent various forms of inequality arising from disparities in wealth, resources, opportunities, and talents. Circumstances concurring, modest gains in both liberty and equality may be achievable without trade-off; but when the realization of either value is pursued more insistently, there comes a point at which one has to be paid for with the other. ‘It is an uncomfortable situation’, Williams concludes already in his influential early essay on the idea of equality, ‘but the discomfort is just that of genuine political thought’ (1973d, 249).

Dworkin believes we can do better. He sees reason to hope that we might eliminate the tension between the concepts of liberty and equality. His leverage point is the idea that whether liberty and equality conflict ‘depends on how we conceive these abstract values’ (2001a, 83).⁴ This seemingly trite observation marks a crucial step: it transposes what might otherwise have been a metaphysical debate about the nature of liberty and equality into a debate about conceptual ethics. As long as liberty is conceived as ‘freedom

³ In reconstructing this debate, which plays out in some of Williams’s most difficult writings, I have benefited from extensive discussions with Damian Cueni, who himself offers a reconstruction of the debate from the perspective of legal theory (Cueni manuscript-a). On his reconstruction, Dworkin is best read as registering the fact that when thinking about political values from the perspective of public decision-makers such as judges, there are real practical pressure on them to speak with one voice and construct a unified and coherent perspective giving everyone their due. But Williams in turn registers the fact that this demand is largely absent when thinking about political values from the perspective of ordinary citizens.

⁴ In Dworkin’s terminology, they are *interpretive* concepts; see Dworkin (1986, 45–86; 2001b).

from the interference of others in doing whatever it is that you might wish to do' (2001a, 84), Dworkin continues, liberty undoubtedly conflicts with equality. But why should we think that we are committed to this way of conceptualizing liberty? The question cannot simply be answered by an inquiry into what liberty *really* is, he observes—we 'can't conduct a DNA analysis of liberty' (2001a, 86). The question has to be *which conception* of liberty we have most reason to accept. And perhaps the most attractive conception of liberty will turn out not to conflict with the most attractive conception of equality after all.

For Dworkin, the fact that two concepts do not conflict is in itself already a reason to prefer them over concepts that conflict: 'integrity among our concepts is itself a value', he maintains, 'so that we have that standing reason for seeking out, for preferring, conceptions of our values that do not conflict' (2001a, 127). Other things being equal, we have a standing reason to pursue what he calls conceptual 'integrity'. In the terminology eked out in the previous chapter, conceptual integrity amounts to the type of coherence between value concepts that we called *congruence*: two concepts are congruent if and to the extent that the realization or instantiation of one concept does not come at the expense of the realization or instantiation of the other concept.⁵

Dworkin therefore proposes to iron out the conceptual tension between liberty and equality by composing a definition of the concept of liberty that renders it systematically congruous with the concept of equality. If Berlin reached the conclusion that liberty and equality irremediably conflict, Dworkin believes, it was because Berlin equated liberty with freedom from interference in doing what one wants.⁶ In fact, however, liberty should be interpreted as a *political* rather than personal value: it is 'that part of your freedom that government would do wrong to constrain' (2011, 4).⁷ As a political value, liberty should not be understood as freedom from interference in doing what one wants, but in terms of *rights* distributed according to a political principle of equality—in other words, liberty

⁵ This is not to deny that Dworkin also sees ulterior reasons to strive for conceptual integrity, such as the hope thereby to reduce political and legal conflict: see Dworkin (2000, 120–83; 2001a, b).

⁶ See Dworkin (2011, 367).

⁷ As Dworkin also puts it: someone's liberty is 'the area of his freedom that a political community cannot take away without injuring him in a special way: compromising his dignity by denying him equal concern or an essential feature of responsibility for his own life' (2011, 366).

should be *rightful* freedom.⁸ That does not presuppose a right *to* freedom; Dworkin argues instead ‘for rights to liberty that rest on different bases’ (2011, 4), such as rights to ethical independence, to free speech, and to due process of law.⁹

If we conceive of the political value of liberty in terms of equally distributed rights to liberty, this ‘rules out genuine conflict with the conception of equality … because the two conceptions are thoroughly integrated’ (Dworkin 2011, 4). The realization of equality may of course still entrain a loss in freedom. But not every loss in freedom will be a loss in liberty. A loss in freedom will only count as a loss in liberty where there is a *claim* to liberty, and there can only be a claim to liberty, on Dworkin’s account, where that claim can be grounded in a *right*. It follows that as long as rights are equally distributed, liberty must itself be equally distributed, and liberty and equality can no longer conflict. As a result, ‘[t]he alleged conflict between liberty and equality disappears’ (Dworkin 2011, 4). This yields an extremely neat and tidy account that immunizes the concepts of liberty and equality against conflict. It achieves this by defining the concept of liberty in terms that effectively guarantee its congruence with the concept of equality.

Williams, on the other hand, has strong reservations about the pursuit of conceptual ‘integrity’—especially in the case of the political value concepts that Dworkin is keenest to reconcile. While Dworkin encourages us to *aim* for conceptual integrity and just see whether we can or cannot construct concepts that achieve it,¹⁰ Williams does not think that there is a totally tensionless conceptual apparatus waiting to be discovered in this way.

Williams’s scepticism towards the attempt to achieve conceptual integrity is grounded, at the most immediate and still somewhat question-begging level, in his acceptance of Berlin’s *value pluralism*: the thesis that there is a plurality of irreducibly distinct and incommensurable values that are bound to end up pulling in competing directions when pursued in concert, not merely because time is short or the world recalcitrant, but because the values themselves inherently conflict.¹¹ As Berlin puts it in

⁸ See Dworkin (2000, 120–83; 2001a, b).

⁹ For Dworkin’s elaboration of what these rights amount to, see Dworkin (2011, 368–74).

¹⁰ See Dworkin et al. (2001, 127).

¹¹ Stocker (1990) offers a detailed elaboration of the pluralist thesis that is informed by Aristotle; Kekes (1993) offers one informed by Berlin and Michael Oakeshott. Here I rely particularly on Williams’s

his best-known essay: ‘We are faced with choices between ends equally ultimate, and claims equally absolute, the realisation of some of which must inevitably involve the sacrifice of others’, which is why ‘the possibility of conflict—and of tragedy—can never wholly be eliminated from human life, either personal or social’ (2002b, 213–214).

Some pluralists, including notably Rawls, have focused on pluralism at the level of society, where the values of some members of society clash with the values of other members of society.¹² But what Berlin and Williams emphasize is that even if society were far less pluralistic, so that members of society shared roughly the same range of values, there would still be pluralism ‘within the breast’: even the concepts of one and the same person conflict in ways that are not resolvable without loss.¹³ A value conflict is not ‘most typically enacted by a body of single-minded egalitarians confronting a body of equally single-minded libertarians’, Williams notes, ‘but is rather a conflict which one person, equipped with a more generous range of human values, could find enacted in himself’ (1981a, 73). Of course, social pluralism and pluralism within the breast are not unrelated, since the latter is in many instances an expression of the former: the conflict enacted within one person may itself reflect the fact that the person’s conceptual repertoire is the accumulated historical deposit of different social influences—be it different groups within a society, different stages in the history of that society, or even different societies.

The cardinal claim of pluralism, however, is that these values, at whichever level they are expressed, are such that all the things they pick out as being of value cannot ‘ultimately be united into a harmonious whole without loss’ (B. Williams 2013, xxxv). The ideal situation in which, as P. F. Strawson put it, ‘every god is given his due and conflict is avoided by careful arrangement and proper subordination of part to part’ (2008b, 30) is

elaboration of pluralism. For an exploration of pluralism’s implications for politics, see Galston (2002, 2005).

¹² ‘No society can include within itself all forms of life’, Rawls writes—‘there is no social world without loss: that is, no social world that does not exclude some ways of life that realize in special ways certain fundamental values’ (1993, 197).

¹³ See Berlin (2013d, 12) as well as Berlin and Williams (1994). Berlin’s animating concern in adverting to pluralism was to cast doubt on the feasibility, in principle, of realizing utopian social arrangements or creating a perfect state (2013a, 48–50; 2013d, 14), and to warn against the danger of allowing this utopian ideal to determine one’s moral and political practice (2002b, 212–17); but logically, these political conclusions lie downstream of a structural claim about human value concepts.

a fantasy, and an incoherent one at that. For the pluralist thesis is not just the weak claim ‘that in an imperfect world not all the things we recognise as good are in practice compatible’, but the much stronger claim, which is a claim about our *concepts* before it is a claim about the world in which they are deployed, ‘that we have no coherent conception of a world without loss, that goods conflict by their very nature’ (B. Williams 2013, xxxv). Our value concepts are incongruent, and while the incongruence may be accidental in some cases, it is non-accidental in others, and hence ineliminable as long as we hold on to anything like these values.

But must we hold on to anything like these inevitably conflicting values? This is the question raised by the radically revisionist conceptual ethics of someone like Nietzsche, and, in a less radical spirit, it is also Dworkin’s response to this pluralism-based worry. Dworkin does not deny that the conceptions held up by pluralists in fact conflict; what he questions is the *authority* of those conceptions. Pluralism, Dworkin complains, ‘is too often cited as a kind of excuse for not confronting the most fundamental substantive issues’, in particular ‘the hard work of actually trying to identify the right conceptions of the values in question’ (2001a, 124–25):

the argument necessary to defend pluralism … must show, in the case of each of the values it takes to be in some kind of conceptual conflict with one another, why the understanding of that value that produces the conflict is the most appropriate one. (2001a, 90)

Values are something that *we construct*, Dworkin reminds us, and finding the most attractive conceptions is not a matter of ‘excavating the shared meanings of words’ or of making ‘anything like a scientific discovery about the true nature of reality’ (Dworkin et al. 2001, 126).

As Dworkin’s claim that we should identify the most attractive *conception* of the concepts we have already brings out, however, his approach is much less radically revisionist than that of someone like Nietzsche: Dworkin does not go so far as to call into question the very use of anything like liberal and egalitarian concepts. Instead, he deploys the concept–conception distinction as a distinction between different levels of abstraction within the concepts already in use: a *concept*, for Dworkin, is abstractly characterizable in terms of what different interpretations of that concept tend to share, while ‘the controversy latent in this abstraction is identified and taken up’ at a more fine-grained level by distinguishing competing *conceptions* of the concept. This yields what

Dworkin envisions as a ‘treelike structure’, the trunk being the concept—what people ‘by and large agree about’—and the branches being the conceptions—the ‘more concrete refinements’ of that concept with respect to which people differ (1986, 70–71). Dworkin offers *courtesy* as an example: people might agree, at an abstract level, that courtesy is a matter of respect, while at the same time disagreeing over what exactly that form of respect requires.

For Dworkin, then, the question is not whether the concepts we inherited conflict, but whether the most attractive conceptions of them that we could be using do. He does not question our continued use of the concepts of liberty and equality. But he is less impressed by the constraints that existing ways of thinking exert on us than by the degree of freedom they leave us to revise our conceptions of our values. Even if pluralists are rights about the conceptual apparatus we inherited, we remain free to try and construct values that do not conflict, and it would be premature to accept the persistent possibility of conflict before every effort had been made to forestall it. As he insists: ‘We shouldn’t buy failure in advance: we should aim at integrity in an optimistic spirit’ (Dworkin et al. 2001, 127).

Yet Williams’s pessimism with regard to our prospects of arriving at such conceptual integrity derives, at a deeper level, from the expectation that any viable set of values will include at least *some* values that necessarily conflict. Like Berlin, he thinks that there are some value concepts that human societies are bound to cultivate in some form—not because these are given to us by God or implanted in us by nature, nor because there is some antecedent structure in reality that these concepts are bound to reflect, but because the conjunction of certain facts about human beings and their environment systematically gives rise to the same kinds of practical problems calling for the same kinds of solutions, which notably include the development of the same kinds of concepts.

Hobbes and Hume may be the best-known exponents of the idea that certain concepts are rendered necessary by contingent facts about us and the world we live in.¹⁴ But the same idea is still very much alive in the twentieth century.¹⁵ It underpins Berlin’s conviction that values vary only within the bounds of a ‘human horizon’, i.e. that there is

¹⁴ Pettit (2008), Cohen (2008), Sagar (2018), and Queloz (2021b) notably foreground this aspect.

¹⁵ See Hall (2020) for a discussion of this idea in the work of Berlin, Williams, and Hampshire. Other exponents of this idea include Edward Craig (1990, 2007) and Miranda Fricker (2007, forthcoming).

a ‘minimum of moral values accepted by all men without which human societies would disintegrate’, (2015, 206). In this respect, Berlin self-consciously echoes his close friend Herbert Hart, who, in *The Concept of Law*, suggests that we can make certain generalizations about what kinds of practical needs will arise in the kinds of environments that humans inhabit, and ‘as long as these hold good, there are certain rules of conduct which any social organization must contain if it is to be viable’ (2012, 192–93). In a similar vein, Stuart Hampshire remarks that if ‘the underlying structure of moral distinctions has no supernatural source, it must be recognized by rational inquiry as having its origin ... in constant human needs and interests’ (1983, 128). We can then look for bounds on the variability of moral, political, and legal concepts, imposed by practical demands faced by human beings nearly everywhere in virtue of their sharing certain very basic concerns, such as the concern to avoid violent conflict with others, to find out about the dangers and affordances of their environment, to secure the resources they need to survive, and to foster conditions enabling cooperation. Williams himself puts a version of this idea to work in *Truth and Truthfulness* (2002, 126) to argue that any human society needs to value the truth for its own sake in order to effectively gain and share information.

When the pluralist claim that some value concepts inevitably conflict is combined with the further claim that some of these are concepts that human societies cannot do without, an argument finally emerges for why the quest for total conceptual integrity is unlikely to succeed. Berlin pithily alludes to both claims in his contention that ‘collisions of values are of the essence of what they are and what we are’ (2013d, 13).

For pluralists like Berlin and Williams, Dworkin’s attempt to inoculate the concepts of liberty and equality against conflict appears as a manifestation of the deep-seated human desire to eliminate tragic value conflicts, i.e. painful conflicts of *right* against *right* that cannot be resolved without loss.¹⁶ It is another instance of the ancient aspiration to use

¹⁶ This conception of tragedy as a conflict of *right* and *right* is associated with Hegel, who writes in his *Aesthetics* that the ‘original essence of tragedy’ consists in a conflict in which ‘each of the opposed sides, if taken by itself, has *justification*; while each can establish the true and positive content of its own aim and character only by denying and infringing the equally justified power of the other’, thereby becoming ‘nevertheless involved in *guilt*’ (1975, II, 1196). See Robert Williams (2012, 120–42) for a discussion of this conception of tragedy, and Magee and Williams (1971, 162–165) for a discussion of its relevance to philosophy.

theorized conceptual constructions as shelters from luck, *techne* as a remedy to *tyche*.¹⁷ Various conceptual edifices—from theological systems through Stoicism to Kantianism and utilitarianism—have been erected over the ages to rationalize away value conflicts when they arose, or to prevent them from arising in the first place. A prime example is the Kantian doctrine that the claims of prudence, in a specially capacious sense of the word, are silenced when they conflict with the claims of morality.¹⁸ But similar dissolutions of conflicts can be achieved by devising lexical priority rules, or reducing values to a common currency of comparison, or arguing that conflicts evaporate under reflection once some obligations are revealed to be merely apparent, or at most *prima facie* obligations.¹⁹ If pluralists are right that concepts we cannot do without carry the ‘permanent possibility of conflict’ because that possibility is ‘implicit in the structure of these concepts as values’ (Williams 2001a, 95), however, the prospects for the pursuit of conceptual integrity across the board are dim.

Alongside this pluralism-based objection to Dworkin’s project, Williams then outlines a second, more fundamental objection that leads him to be critical of the very ambition to achieve conceptual integrity in this connection. He suggests that we discover a real *need* to keep open the rift between the concepts of liberty and equality that Dworkin offers to patch up with his tidy conceptual construction once we reflect on the central *concerns* that render the concepts of liberty and equality important for us in the first place: what is it that we fundamentally care about in these connections that leads us to have a use for anything like these concepts? Our concepts may channel our attention and shape our sense of saliency in various ways. But the merits of a proposed concept ultimately have to be judged on the basis of a prior understanding of the concerns—the needs,

¹⁷ See Berlin (2013b, 196; 2013c, 26–28; 2014a, 25; 2014b, 99–100) and Williams (1981c, 20; 1995e; 2011, ch. 1). See also Nussbaum (2001), Queloz (2022b), and Queloz and van Ackeren (manuscript) for further discussion of this aspiration.

¹⁸ For an analysis of this Kantian idea, see Bader (2015).

¹⁹ See Berlin (2002a, 291–292; 2014b, 61–62, 70–72). W. D. Ross introduced the notion of a *prima facie* obligation to register a claim or responsibility that is more than a *seeming* obligation but stops short of constituting an obligation proper (1930, 20). Even if we conclude that this notion of a *prima facie* obligation cannot ultimately do any real work within the constraints of a strictly deontological framework, it nonetheless constitutes an attempt to register something important: that an obligation, even when outweighed by another obligation, can leave a residual claim that gives the agent cause for regret.

interests, desires, projects, aims, and aspirations—that this concept is to help us to meet. Though our concerns may be mediated and focused by our concepts, it is fundamentally the direction of those concerns that determines what we care about, what is important or relevant to us (the Latin *concernere* means ‘be relevant to’, from *con-*, which expresses intensive force, and *cernere*, which means ‘sift, discern’). Indeed, Williams suggests that we only really grasp why *liberty* and *equality* are two different concepts, and not just two words for the same concept, once we relate them to the concerns underlying them.²⁰

Of course, there are many concerns that concepts such as *liberty* and *equality* tie in with; but Williams suggests that we can to some extent cut through that complexity, because ‘associated with each such value concept there is a kind of schema, a very bare outline of what our central concern is’ (2001a, 92). In the case of the concept of equality, for instance, a basic human concern that might lie at its root is the concern *to receive what one is due* when desiring something one lacks.²¹ But our use of the concept of liberty is animated by a different and utterly basic concern, namely the universal human concern *to be unobstructed in doing what one wants*—in particular, unobstructed by humanly imposed coercion. He labels this the concern for ‘primitive freedom’. ‘Why should human beings in general be concerned with some value of that form?’, Williams asks, and answers his own question with another: ‘What view would one have to take of one’s desires and projects and other values if there were never even a question of its being something to be resented and resisted if others aimed to frustrate them?’ (2005c, 93). His point is that anything recognizable as human agency is bound to involve some concern along the lines of the concern for primitive freedom, because pursuing any concerns at all must already involve being concerned to be unobstructed by others in doing so. A human agent who pursued concerns without caring at all about whether they were frustrated by others simply would not be intelligible to us as genuinely pursuing

²⁰ See Williams (2001a, 92).

²¹ This is how Cueni (manuscript-a) helpfully glosses Williams’s all too brief discussion of the concern at the root of the concept of equality. In the New York debate, Williams shifts from speaking of a concern for ‘equality’ to speaking more broadly of a concern for ‘justice’, and then suggests that the concern at the root of the notion of justice is that of ‘giving each person what he or she should have, or some similar idea (2001a, 92). But his more detailed responses to Dworkin—i.e. Williams (2005a, c)—are cast in terms of ‘liberty’ and ‘equality’. For a detailed account of the concerns that underlie and drive the development of our notion of equality, see Sagar (forthcoming).

those concerns.

By itself, of course, that schema is ‘too bare, indeed too schematic’, Williams acknowledges; it ‘needs, and will have indeed received, an associated social, historical, and cultural elaboration’ (2001a, 93). For a start, this concern for primitive freedom is not yet a political concern; it is a personal concern for freedom in action, antecedent to and intelligible independently of the advent of the political. But it is through the elaboration of a political concept of liberty, Williams contends, that this pre-political concern for freedom finds a suitable political expression (a suggestion I unpack in Chapter 9).

Hence, what we really want to know, and what Dworkin ignores in his pursuit of conceptual integrity, is to what extent a given elaboration of a concept serves, or fails to serve, the concerns we now have. For whatever exactly the concerns that basically go with the values of liberty and equality now go *to*, we cannot *redirect* those concerns ‘simply nominalistically, by redefining a word’, Williams insists, because ‘an interest in producing a more coherent body of law is not by itself going to stop the concern going to what the concern goes to’ (2001a, 94). If Dworkin’s proposed concepts fail to tie in with the concerns that give us reason to think in terms of *liberty* and *equality* in the first place, we have reason *not* to adopt those concepts, because they would deflect attention away from the satisfaction of our most basic concerns in these connections. It is simply no good securing conceptual integrity between two concepts if it comes at the cost of severing the ties to the central human concerns that led us to care about these concepts in the first place.

This concern-based objection yields an argument that promises to do more with less than the pluralism-based objection: without requiring sweeping assumptions about what values societies are bound to possess, it calls into question the very ambition to iron out tensions in our conceptual apparatus. It suggests that it is a fundamental mistake to look only at conflicts or coherence between *concepts*. We have to look also at the human *concerns* underlying the concepts—at *what we care about* in these connections. These concerns are in significant respects independent of our concepts, so that constructing more coherent or otherwise more theoretically virtuous concepts is not by itself going to redirect the concerns underlying the concepts we inherited. And if it is by tying in with these concerns that the concepts of liberty and equality become concepts worth using in

the first place, we are clearly not helped by securing coherence or ‘integrity’ between these concepts if they thereby lose their connection to the concerns animating their use.

Williams’s fundamental worry is thus a concern-based worry about the advisability of striving for conceptual integrity. Even if we *could* achieve total integrity or congruence between our concepts, we would be ill-advised to do so if it untethered our concepts from our concerns. This suggests that eliminating tensions and cultivating theoretical virtues in our concepts is not necessarily advisable. Conflicting concepts can be the better concepts in virtue of serving our concerns better.

5.2 Lessons from the Dworkin–Williams Debate

I believe that a number of important lessons can be extrapolated from the Dworkin–Williams debate. Some of these bear specifically on how our political concepts should deal with the conflictual nature of politics—an issue we shall return to in Chapter 9. But the insights to be gained from the debate also have ramifications beyond what they tell us specifically about the conflict between *liberty* and *equality*. They point towards an approach to the authority of concepts that ties concepts back to the concerns of concept-users.

What renders the Dworkin–Williams debate particularly interesting in connection with a discussion of conceptual authority is, first, that Dworkin himself frames the issue in terms of which conceptions we have most reason to accept; second, that he points to the conceptual integrity of his proposed conceptions as a reason to accept them; third, that Williams draws on the pluralist idea that our values ineliminably conflict to question the attainability of this kind of harmony in our conceptual apparatus; and fourth, that he introduces the important idea that our concepts should serve our concerns, and derives from it the conclusion that if these concerns conflict, this should lead us to question not just the feasibility, but the very advisability of eliminating tensions between our concepts. This offers a sharp contrast not just to Dworkin’s pursuit of conceptual ‘integrity’, but to the prevalent view of conceptual tensions, which is that they should always be eliminated. On Valerie Tiberius’s view, for example, conflict between our value concepts constitutes a form of ‘ill-being’ that we should seek to overcome as we pursue well-being through the

fulfilment of our values.²²

I propose to extrapolate three main lessons from the Dworkin–Williams debate. The first and fundamental one is that the merits of a proposed concept ultimately have to be judged on the basis of a prior understanding of the life that this concept is to help us to lead, and this requires us to understand the various *concerns* that motivate our use of certain concepts in the first place. The authority of concepts does not trickle down from the theoretical virtues of the tidy conceptual structure into which they interlock. It wells up from the concerns underlying people's use of these concepts.

By 'concerns', I mean more settled concerns, not momentary whims and impulses. Moreover, as the examples of the political concepts of liberty and equality already intimate, what lends a concept authority for me need not be a *selfish* concern to get something for myself. Nor need it be a *self-centred* concern, in the sense in which an artist's concern to create great art—not for himself, but for the world—might be said to remain self-centred if it has to be *him* who creates it.²³ The relevant concern can be an altruistic and selfless concern for reform, or for the advancement of liberty or equality. It need not even be, in the first instance, *my* concern at all; it can be a shared concern, the project of a group or community. Even if our concerns are, at some level, what infuses the world with significance for us, what we will be concerned *with* as a result is the world, not ourselves.

At the same time, it is not enough for a concept simply to serve *some* concern or other. If the concerns being furthered by a concept are those of hostile parties, my use of the concept works against my own interests, and realizing this would hardly strengthen my confidence in that concept. The term 'ideological', in one of its uses, captures just this sense in which the use of a concept might be in the interest of one group while going against the interests of another group. So the concern in question does need to be a concern I acknowledge as authoritative, a concern that, in this sense, I *identify with*. It needs to be a concern *for* me.

Of course, how closely I identify with a concern might change over time; at the limit, my identification with a particular concern might last only for the duration of a particular project, in which case the concept I have most reason to use in relation to this concern

²² See Tiberius (2018, 34–35).

²³ See Williams (1981e, 13).

will also only be authoritative for me as I pursue that project. But the guiding idea is that, however external and unrelated to me a certain concern might be *in itself*, my coming to *identify* with it makes it a concern *for* me, and in virtue of its being a concern for me, I have a *pro tanto* reason to prefer concepts that help the satisfaction of that concern over concepts that hinder it.

In some cases, the concerns will be concerns I *deeply* identify with. The concerns will be, for me, what Williams calls ‘ground projects’ or ‘propelling concerns’, which is to say concerns that propel one to live. These propelling concerns, he observes, ‘do not have to be even very evident to consciousness, let alone grand or large . . . the propelling concerns may be of a relatively everyday kind such as certainly provide the ground of many sorts of happiness’ (1981e, 12).

But equally, some of the concerns I identify with may have been formed in ways that make them questionable bases for conceptual authority. What of concerns that have been instilled in the oppressed by the oppressors so as better to oppress them?

The basic thought that conceptual authority wells up from the concerns we identify with accordingly requires further qualification to rule out concerns that merely reflect easily corrected misapprehensions, or concerns that one has been deceived into forming against one’s own interests, but that would immediately go dead on one the moment one realised how one came by them. We might say that the concerns we should look to in answering the authority question should be concerns we identify with *and would still endorse after critical reflection on how we came by them*.²⁴

This also helpfully underscores that one’s identification with and endorsement of a concern is not static, but dynamic: it is responsive to reflection and to changes in one’s sense of what is possible. Indeed, the maelstrom of our inchoate wishes arguably only settles into fully formed concerns in response to social pressures, such as the demands that others make on us to present ourselves one way or another, declare what we want, and take a stand.²⁵ Even once settled, however, a concern can be unsettled again, called into question by the critical prodding of others, or by the experience of an uncomfortable

²⁴ I ignore here the relevant and well-known difficulties involved in specifying what exactly is allowed to go into such a process of reflection, and in excluding problematically self-validating forms of that process. I return to these issues in Chapter 7.

²⁵ See Williams (2002, 191–98), Fricker (2007, 52–53), and Pettit (forthcoming).

tension between that concern and the rest of one's concerns. One's identification with a concern is thus responsive to critical reflection on it in light of one's other concerns.

The suggestion that concepts should tie in with the concerns that underlie them points to a more agent-centred way of thinking about conceptual authority than any of the conceptions we have encountered thus far. It indicates a conception on which conceptual authority ultimately derives from the way a concept makes contact with our concerns. This offers an attractive alternative to more established ways of appraising concepts. It differs not just from approaches that look at concepts in isolation and ask whether they display certain theoretical virtues, but also from approaches that focus on a concept's referential relation to the world and ask whether it is determinate and joint-carving,²⁶ and from approaches that look at a concept's inferential relations to other concepts and ask whether these inferences are reliable.²⁷ Like Nietzsche before him, Williams underscores the importance of looking beyond concept-to-world and concept-to-concept relations to try and appreciate how a concept with certain referential and inferential characteristics relates to the human concerns that animate its use.

The second lesson we can extrapolate from the Dworkin–Williams debate is that whatever concerns concept-users genuinely have cannot be eliminated or redirected at the drop of a definition. This identifies an important source of obstance or recalcitrance faced by attempts to improve our conceptual apparatus by conceptual innovation, redefinition, or stipulation: the difficulty is not just that of securing the uptake of newly minted concepts; it is that the underlying concerns cannot be redirected by fiat. This is not to say that our concerns are insensitive to the concepts we use—the human concern for beauty will be given a different direction and expression if focused by the concepts of Babylonian painters than if focused by those of *Der Blaue Reiter*. But when the tension between two concepts is reflected in a corresponding tension between two concerns, analysing away the tension at a conceptual level will do no more to resolve the tension at

²⁶ This referential dimension is foregrounded in, e.g., Hirsch (1993, 2013), Sider (2011), Cappelen (2013), Sawyer (2020a, c), and the essays in Campbell, O'Rourke, and Slater (2011).

²⁷ This inferential dimension is foregrounded in, e.g., Brandom (1994, 2000, 2001), Cohen (2016), Peregrin (2014), Greenough (2020), Fraser (2018), Mühlebach (2021, 2022, Forthcoming), Scharp (2013, 2020, 2021), and Jorem and Löhr (2022). Tristram McPherson and David Plunkett also focus on what they call 'the *unreliable inference* danger' (2020, 283).

the level of our concerns than painting over a crack does to close the crack. The tension will need remedying not at the level of our concepts, but at the level of our social arrangements. New concepts fastidiously engineered to cohere might successfully replace their predecessors without doing anything to attenuate the conflict between the underlying concerns—they might even leave people worse off by rendering them conceptually ill-equipped to recognize at what price these concerns are being pursued. Real costs can be incurred even when one is conceptually blind to them. Without a corresponding redirection of concerns, the revision of concepts cannot create more than an illusion of harmony.

The third lesson, finally, which falls out of the former two, is that if some proposed concept realizes theoretical virtues at the cost of severing its tie to more important concerns, it will not be an improvement. As the example of the concept *person* in Chapter 4 brought out, it is our concerns as mediated by our concepts as they are *before* the engineer's intervention that an engineered concept must make contact with in order to have a claim to authority. When these antecedent concerns are themselves primarily directed at the achievement of theoretical virtues such as determinacy, consistency, or coherence, the concepts engineered to realize such virtues can be authoritative. But in contexts in which we are not primarily concerned to realize these theoretical virtues, concepts engineered to be free of generic theoretical defects such as vagueness or inconsistency will carry *less* authority than the concepts they are meant to replace. Theoretical vices can be practical virtues.

Let me expand on this third point, as it is crucial to motivating the departure from the more orthodox form of non-foundationalism that seeks authority in theoretical virtues.

5.3 The Practical Virtues of Theoretical Vices

To think that more theoretically virtuous concepts must be more authoritative than concepts exhibiting theoretical vices is to model concept appraisal on theory appraisal: in a scientific theory, it is a defect in our concepts if they conflict, and the defect is to be fixed by adjusting our concepts. Likewise, in a linguistic theory aiming to explain and predict competent speakers' intuitions about which sentences of a language are well-formed, it makes sense to expect the theory to work the linguistic intuitions it takes as input into a consistent body of rules which can then be used to adjudicate between

inconsistent intuitions, so that one ends up adjusting conflicting intuitions to make them fit the theory as much as adjusting the theory to make it fit the intuitions.

But a person torn between the demands of two value concepts is not like a drunk polyglot mashing together two languages. While consistency is a constitutive aim in formulating a system of linguistic rules, inquiry into whether concepts and their concomitant reasons have a foot in real concerns may find reasons for reasons on both sides of a conceptual conflict. This will affirm rather than dissolve the conflict, giving us reasons to sustain it and work through it.

Accordingly, while the drunk polyglot merely has to discipline himself into sticking to one of several mutually exclusive sets of linguistic rules, the person experiencing a conflict of values may have good reason to ride out the tension. She may look for ways to absorb or accommodate the tension, but only to the extent that these allow her to remain responsive to both of the concerns that give rise to the tension.

In the case of the thick normative concepts articulating moral or political judgements, it is thus much less clear that a tension between two concepts is necessarily a defect at all: tensions between concepts can express real tensions between the things we care about, and when this is the case, we may prefer our concepts to mirror these tensions. A lack of consistency does not necessarily imply a lack of conceptual authority, just as consistency does not by itself guarantee conceptual authority. In fact, if Williams is right in his critique of Dworkin, rendering the concepts of liberty and equality consistent actually constitutes a deterioration rather than an amelioration, because it blinds concept-users to a real conflict between their concerns.²⁸ Insofar as we would be ill-served by a concept that blinded us to this real conflict, we therefore have reason *not* to adopt the concepts immunized against conflict that Dworkin advocates. They would put us out of touch with our concerns and the conflicts between them. Our concepts of liberty and equality *should* conflict, because our *concerns* do.

This suggests that theoretical virtue is not necessarily something we should strive for across all our concepts. There are conceptual virtues beyond theoretical virtues. We should look to the concerns underlying the use of concepts before we look to the inherent

²⁸ The argument for this conclusion turns on a complex political account of why the concept of liberty best serves our concerns by retaining its inconsistency with the concept of equality (Williams 2001a, 2005a, c), which I come to in Chapter 9.

defects of concepts, and understand what *counts* as a defect *in relation to* those concerns. Inconsistent concepts can be the better concepts in virtue of serving our conflicting concerns better. Where our concerns conflict, the correspondingly conflictual character of our concepts will be an expression of truthfulness—a way of being true to the real and conflicting directions of our concerns.

To insist that our moral and political concepts should be subjected to precisely the same kinds of demands for theoretical virtues that we make on the concepts forming the building blocks of our scientific theories is then not necessarily to render our thinking in some incontestable sense more rational. It might instead amount to adopting a radically revisionary and highly contentious stance in conceptual ethics: for insofar as the moral and political concepts we have conflict *for good reasons* rather than for lack of trying to construct more congruent concepts, imposing the requirements we make on theories on the concepts articulating our moral and political outlook is tantamount to advocating a radical reconfiguration of that outlook, and hardly one that is incontestably more rational.²⁹

The upshot is that when concepts conflict, we need to understand *why* we have conflicting concepts. If the conceptual tension is the result of committing an intellectual error within an enterprise aiming at systematic unity, the conflict should indeed be rationalized away in the name of consistency, coherence, and congruence. If, however, the conceptual tension is rooted in an underlying conflict between concerns we identify with and would endorse upon critical reflection, a more complex response is called for. A judgement is required on what kind of balance to strike between the relevant concerns.

What can inform and guide that judgement is a deeper understanding of where the concerns come from and how they come into conflict: what are the historical, social, and psychological dynamics that produced this conflict? The conflict might be the result of having accumulated, within one society, the legacies of different historical periods or the influences of different societies; or it might reflect changes in external circumstances which now bring into tension with each other concerns that used to coexist without conflict; or, as in the case of *liberty* and *equality*, it might be grounded in enduring features of human psychology that make it well-nigh impossible to envision a human society in which conflicts along these lines never arose. In some cases, understanding what

²⁹ See Williams (2006h, 162–4).

produces a tension between two concepts may lead one to withdraw one's confidence from one of the concepts involved, thereby eliminating the tension. But in other cases, such understanding will only affirm the tension by showing that the conflicting concepts each have a foot in legitimate concerns that decidedly pull in different directions. There are ways to rationalize conceptual conflicts without rationalizing them away.

None of this is to deny that, in a concrete situation in which two of our value concepts conflict, we should reduce the tension between them as far as we can. But this will be conflict-reduction on a piecemeal basis, where we attempt to find a rational resolution to conflicts *when they arise* instead of trying to *forestall* their arising in the first place by immunizing our conceptual apparatus against conflict.

The fact that a conceptual apparatus persistently gives rise to tensions and value conflicts need not even be seen as something to be regretted: it can itself be seen as a strength, not just of the person who manages to live with the tensions, but of the conceptual apparatus itself. Its tensions can be an expression of the scale and scope of its aspirations, of its richness and diversity, and of its sophisticated capacity to discern value along multiple dimensions and in different things at the same time.

Moving towards a different set of concepts comprising fewer, less diverse, and less demanding values would reduce conceptual tensions, and, circumstances concurring, it might permit the joint realization of those values without loss. But this does not mean that there would be no loss. Yes, life would *seem* easier for those living by an outlook that had been immunized against conflict. But to immunize our conceptual apparatus against conflict would be to ignore the force of conflicting concerns that are really there. In order for these concerns to be adequately acknowledged in our deliberations, the tensions between those concerns must be reflected in the tensions between our concepts. Resisting the demand to tidy up our concepts then expresses a determination to truthfully confront the conflicting demands we face, and the real losses incurred by satisfying one demand at the expense of another, instead of escaping into comforting fantasy.

Of course, this might be taken to encourage going further in the elimination of tensions by eliminating not just the conflicting concepts, but the conflicting concerns. One could abandon the ethical, political, or aesthetic concerns that produce the tensions and the resulting sense of loss in the first place. From the perspective of a tension-ridden

outlook, however, the transcendence of conflict made possible by such a transition to tension-free concerns would still involve a significant loss: peace of mind would be bought at the price of indifference, requiring withdrawal from the world into the kind of apathic aloofness that experiences no loss only because it does not care sufficiently about enough things. The pursuit of the perfectly tension-free life thus risks issuing in something which, while tension-free, is rather less of a life. Those fully identified with the tension-free concerns may not share that sense of loss. But this merely means that the loss would include the loss of the sense of loss—which only makes it more of a loss.³⁰

Thus, understanding why concepts conflict does not necessarily dissolve the conflict between them, or tell us which concept to prioritize. But it can help us to grasp the place and urgency of the aspiration to achieve theoretical virtues such as consistency, coherence, and congruence. The pursuit of theoretical virtues has a place, and is sometimes of the first importance—but not every kind of concept obeys the same constraints, and some concerns are better served by concepts that are less consistent, coherent, or congruent.

These qualifications also hold for other theoretical virtues.³¹ Vaguer concepts, for example, are not necessarily worse concepts, though philosophers' tidy-mindedness can make it hard to recognize that fact. Gottlob Frege even questioned whether concepts with blurred boundaries were *bona fide* concepts at all: comparing concepts to areas, he argued that since an area without clear boundaries was not really an area, concepts without sharp boundaries were not *bona fide* concepts. Yet this betrays a picture on which the rational use of concepts must in no way rely on normality, i.e. on contingent but nonetheless typically reliable regularities in how human beings and the world in which they deploy their concepts are disposed to behave.³² As Wittgenstein observed in response to Frege,

³⁰ For germane remarks on the idea that conflicts of values cannot be transcended without loss of the sense of loss, see Williams (1981a, 80) as well as Williams (1973c, 177; 2001a); for a sustained discussion of concept loss and the loss of the capacity to recognize concept loss, see Diamond (1988).

³¹ In a detailed case study of the concept of divorce, for example, Elizabeth Anderson argues that a *less* fruitful conception of divorce can be preferable to a more fruitful conception for the purposes of a certain research programme (2004, 20). On the merits of less determinate concepts for social critique and political movements, see also Santarelli (2022).

³² This is what Warren Goldfarb calls the ‘demand for fixity’, which Frege shared with the early Wittgenstein, and which Goldfarb thinks expresses ‘an incorrect picture of rationality’ (1997, 79).

we often identify areas without drawing clear boundaries, merely indicating them through pointing gestures.³³ Similarly, concepts without clear boundaries can be *bona fide* concepts, and to deny this, Wittgenstein remarked, would ‘be like saying that the light of my reading lamp is no real light at all because it has no sharp boundary’ (1958, 27).

A vaguer concept will sometimes serve our concerns better than a more precise one. For example, it is not unheard of for legislators to intentionally *blur* a concept that proved too precise: the U.S. Congress once found that its Social Services Administration disability program was better served by a concept of disability that was vaguer than the one previously in use, because this vagueness was precisely what gave low-ranking administrators the discretion they needed to do justice to cases that the sharper definition categorically excluded.³⁴ When Orwell deplores the ‘sheer cloudy vagueness’ of political language, his complaint is ultimately an epistemic one: that vagueness is obfuscating, clouding our grasp of the facts on the ground and thereby our political judgement.³⁵ But vagueness, especially within large institutional structures, can also stand in the service of accuracy and be an expression of truthfulness: it can serve to acknowledge that others, precisely because they have a better grasp of the facts on the ground, are epistemically better placed than we are to determine what exactly should or should not fall under a concept.

Vague concepts can thus be exactly what we need given our concerns. In recent years, a number of legal philosophers, computational linguists, and game theorists have extolled ‘the value of vagueness’ across a wide variety of cases.³⁶ As David Lanius (2021) concludes after a comprehensive survey of that literature, these arguments typically do not show that vagueness is itself inherently valuable. But the value of a vague concept need not be the value of vagueness; its value can lie, rather, in the value of using a concept that better serves one’s concerns in virtue of being vague.

Moreover, vague concepts can achieve their own form of orderliness through

³³ See Wittgenstein (2009, §71).

³⁴ See Zacka (2017, 53–55) and Mashaw (1983, 52–53).

³⁵ See Orwell (2008).

³⁶ Endicott (2000), Soames (2011), Asgeirsson (2015, 2020), Lanius (2019), and Chadha-Sridhar (2021) defend the value of vagueness in the law; game-theoretic rationales for vagueness are identified by De Jaegher (2003), van Deemter (2010), and De Jaegher and van Rooij (2011); on the experimental evidence for the value of vagueness, see Green and van Deemter (2019).

coordinated vagueness. David Lewis points out that one vague concept can neatly interlock with another vague concept that is vague in an exactly complementary way:

It often happens that two vague concepts are vague in a coordinated way: firmly connected to each other, if to nothing else. The border between blue and green is not well fixed, so ‘blue’ and ‘green’ are both vague. But their relation to each other is fixed: one begins where the other leaves off, with no gap and no overlap. (1973, 92)

Vagueness or indeterminacy can also prove to be conceptual virtues when it comes to building broad coalitions around what is perceived as a common grievance. Precise definitions and fully determinate implications then stand in the way of political success by making it harder to conceptualize a grievance as common; the most helpful concepts for politics are often those that can be shared across multiple perspectives precisely because they are vague and indeterminate in their consequences.

Some, like Herbert Marcuse, even went so far as to argue that vague or indeterminate concepts were the better tools for radical critiques of the social order.³⁷ In addition to making it easier to rally around common grievances, vague and indeterminate concepts without clear referents can also be expressive of a deep malaise and a radical dissatisfaction with social arrangements. This subversive moment is lost when the grievances articulated in vague and sweeping terms are reconceptualized in precise and narrowly focused terms that target concrete failures in the operation of the system while taking the system itself for granted. Marcuse’s main example concerns the Hawthorne Works of the Western Electric Company, where the workers’ vaguely articulated protests were recast by industrial sociologists, who sought to make their complaints less vague and indeterminate by transposing them into precise descriptions of concrete episodes: a complaint about dangerous and unsanitary working conditions would be transposed into a complaint about a certain occasion on which the washbowl had some dirt in it. As a result, what began as a collective complaint, inchoately but radically calling into question the system as a whole, was transformed into a loose collection of defanged complaints, individualized and particularized until they each separately aimed at making the system run more effectively instead of calling it into question with one voice. When theoretical virtues are pursued in this way, they risk resulting in what Marcuse calls a ‘repressive

³⁷ See Marcuse (2002, 110–23).

reduction of thought' (2002, 111).

Though one might quibble with the details of Marcuse's analysis, it points to a real worry: re-engineering thought for theoretical virtues may end up assimilating it into the existing social order in a way that takes the sting off, and then, *pace Orwell*, the theoretical virtue of clarity will have become a political vice.³⁸ This worry is amplified by the fact that, as Alexander Prescott-Couch notes, the intermediary actors who tidy up vague and inchoate complaints 'tend to be social elites, and engaging with inchoate speech through their filter might threaten to overwrite or sanitize inchoate voices, particularly those from marginalized communities' (2021, 497–98). There is thus a real danger that if interpretative intermediaries rearticulate vaguely articulated concerns to maximize theoretical virtues, their reconstruction will realize those virtues at the cost of losing any real connection to the original concerns.³⁹

Isolating concepts from the particular concerns with which they tie in and concentrating on the inherent defects of concepts—or what appear to be defects when measured against some ideal of theoretical virtue—thus embodies a strategy that may answer the authority question in special cases, but that cannot hope to do so more widely. Life is not a logic test. The most theoretically virtuous concepts are no good to us if they do not serve the central concerns that animated our use of anything like these concepts to begin with. By treating the respects in which concepts fall short of theoretical virtues as defects, we run the risk of neglecting defects arising from the way in which theoretically virtuous concepts fail to serve the concerns of concept-users. To comprehensively assess the authority of our concepts, and also to give direction to interventions in our conceptual apparatus while ensuring that their results are authoritative, we need to look beyond theoretical virtues, to how concepts tie in with the

³⁸ See Orwell (2008). For a nuanced reevaluation of the risks and benefits of combating unclear language in politics, see Gibbons (manuscript).

³⁹ Although, as Prescott-Couch (2021, 508–16) goes on to argue, not all forms of rational reconstruction by such intermediaries are bound to lose the connection to the original concerns: well-designed ways of dividing the interpretative labour could in principle rearticulate inchoate concerns so as to facilitate a deliberative partnership that is respectful of the original concerns while also rendering them more suitable for inclusion in high-quality public deliberation—for, as Prescott-Couch helpfully highlights, it is the demands of high-quality public deliberation that invite a rearticulation of the vague in the first place (a point I return to in Chapter 9).

concerns of those who use them. These concerns are not always, or even mainly, concerns for theoretical virtue. In relation to the rest of our concerns, even theoretical vices might prove to be practical virtues.

5.4 The Limits of Concerns: Four Problems

The idea that human concepts should be answerable to human concerns is of course not unique to Williams. Several philosophers thinking about concept appraisal have recently floated the related, if somewhat narrower idea that concepts should be evaluated by the lights of our aims or goals.⁴⁰ Such aims or goals are specific types of concerns: those that concept-users consciously pursue in using a concept. Whether on the basis of concerns, aims, or goals, however, few have elaborated on how concept appraisal is to be concretely put into practice or operationalized. And once one does try to operationalize it, real difficulties emerge for attempts to appraise concepts directly by our aims, goals, or concerns.

One difficulty lies in moving from a given concept to the concern or goal that is supposed to help us appraise it. How do we identify the relevant concern? Part of the difficulty is that concerns are not only many and various, but often at variance, and not every concern merits satisfaction. But even before one wades into the politics of conflicting concerns, there is a more basic problem involved in identifying a *relevant* concern to begin with: how does one move from a concept to a concern by which to appraise it? Call this the *concern identification problem*.

Some have argued that the concern identification problem can be overcome by looking for concepts that are inherently tied to a concern. Ingo Brigandt and Esther Rosario, for example, argue that certain scientific concepts have inbuilt epistemic aims:⁴¹

there are cases where a scientific aim can be tied to an individual concept in that this concept is being used by scientists to pursue this aim. For example, while the CLASSICAL GENE concept

⁴⁰ See, e.g., Anderson (1995, 2001), Brigandt (2006, 2010), Haslanger (2012), Burgess and Plunkett (2013b, 1105), Brigandt and Rosario (2020), Nado (2020), and Pinder (2022).

⁴¹ See also Brigandt (2010, 2011, 2012). Brigandt and Rosario (2020) explore how such goal-based appraisal might be extended to socio-political aims. But exactly how one identifies such aims and their relation to the concept at issue does not become as clear as it does in the epistemic case. I therefore concentrate on the epistemic case.

was used for the purpose of predicting (and statistically explaining) phenotypic patterns of inheritance across generations, the MOLECULAR GENE concept serves the aim of causally-mechanistically explaining how inside a cell a gene leads to the formation of its molecular product. Making explicit such an aim tied to a concept's use permits one to account philosophically for the rationality of concept change: a revised definition is an improvement over an earlier definition if the former is empirically more conducive to meeting this aim. (2020, 102)

The idea is 'to view a concept as being used by scientists to pursue a *specific scientific aim*', for it is this aim, which Brigandt (2010) also calls the 'epistemic goal' of a concept, that 'sets the standards for whether one definition of a concept is superior to another definition' (Brigandt and Rosario 2020, 102).

According to this goal-based approach, a concept such as the *classical gene* concept has an epistemic goal insofar as it is used *with a view to* predicting and explaining inheritance patterns. Insofar as the concept's reference and inferential role are not optimally suited to meeting this goal, however, one can distinguish between the inferences a concept *in fact* supports and those it is *intended* to support: while 'a concept's inferential role is the set of inferences and explanations currently supported by the concept', Brigandt writes, 'the concept's epistemic goal is the kinds of inferences and explanations that the concept is intended to support' (2010, 24).

On this goal-based approach, 'the notion of the epistemic goal pursued by a concept's use is ... considered a component of a concept' (Brigandt 2010, 22). The idea is thus not that the *classical gene* concept just happened to be harnessed to pursue a certain epistemic goal while the same concept might have been recruited to different ends; rather, as Alejandro Pérez Carballo helpfully puts it, the epistemic goal is regarded as being *constitutive* of the concept: 'a concept that is not put to use for the purposes of predicting inheritance patterns would not be the classical gene concept' (Pérez Carballo 2020, 305). On this view, the relation of a concept to the concern providing the standard for its appraisal is an *internal* relation: a relation that is intrinsic to the identity of at least one of the *relata*—in this case, the *classical gene* concept.

If we identify the *classical gene* concept as the concept that was intended by its originators to support inferences explaining and predicting patterns of inheritance, therefore, the concept can be appraised according to the extent to which it allows its users

to draw adequate explanatory and predictive inferences—something the concept became increasingly good at doing as it evolved from the *classical gene* concept of William Bateson, which still assumed a one-to-one correspondence between genes and traits, to the *classical gene* concept of T. H. Morgan, which allowed for one gene to affect many traits and one trait to be affected by many genes.⁴² Taking a concept's epistemic goal as a standard of appraisal thus promises to make sense of conceptual amelioration in the context of scientific theory-building. Moreover, it allows us to tailor the standard of appraisal to the concept at issue, alerting us to case-specific desiderata that would have been obscured by uniformly applicable desiderata such as avoiding reference failure or inferential unreliability.

But there are two reasons why, in appraising concepts, one might want to look beyond such inbuilt goals. First, it is not clear that this approach can be generalized, as many concepts do not come with a clear constitutive goal that could guide our appraisal of them. As Brigandt himself admits: 'specific epistemic goals that are particular to a concept may exist only for scientifically central concepts' (2012, 99). There is therefore a *generalizability problem*.

Second, this kind of appraisal remains too internal to the concept and does not allow for enough critical distance. To appraise a concept by its constitutive goal is still to appraise it by its own lights rather than from a more independent perspective. It does not allow one to criticize a concept when that concept is ideally suited to meeting its constitutive goal. Yet the ability to criticize a concept even when it meets its constitutive goal is a clearly crucial desideratum for an account of concept appraisal. Without it, we would be hard-pressed to make sense of paradigm shifts in science or ideology critique in social thought. When the concepts are the mainstays of the old paradigm one seeks to break out of, or when they are ideological concepts promulgated with a view to stabilizing oppressive regimes, taking concepts' constitutive goals as one's basis of appraisal points one in precisely the wrong direction, since the last thing one wants to do is to tailor one's concepts even more closely to those goals. In those cases, a more external basis of appraisal is required, one that makes it possible for concept-users to conclude that even a concept which is optimally designed to meet its constitutive goal is at odds with their own concerns—it is purpose-built to the wrong purpose. To this end, it is crucial that we

⁴² See Weber (2005, 195–96) and Darden (2006, 235–6).

do not always simply buy into the standards encoded in concepts' constitutive goals. In addition to the generalizability problem, relying on built-in concerns thus creates a *critical distance problem*.

One might hope to address these two problems by detaching the standard of concept appraisal from the concept itself, seeking it instead in concerns that are *externally* rather than internally related to concepts. But this reintroduces the identification problem that the focus on constitutive goals was meant to solve, namely the problem of how to work back from a concept to a concern that is relevant to its appraisal.

Even if a relevant concern can be identified, moreover, the concern by itself leaves the concept it calls for underdetermined. Two concept-users with the same concern will be in completely different practical situations if one has useful capacities to draw on—such as the capacity to hear high-pitched sounds or detect subtle smells—that the other lacks; equally, they will be in different practical situations if one is in a congenial environment that abounds with whatever one could wish for while the other is in an arid and hostile environment. By itself, a concern is therefore insufficient to determine what concept we should use. A concern provides no guiding sense of what makes a good concept in relation to that concern, because it fails to give us a sufficient understanding of the dynamics that *mediate* between the concern and the concept, and hence of the practical pressures acting on the concept as result of pursuing the concern. Yet it is in the crucible of these pressures that the concept must prove its worth as a means of meeting the concern. We thus also face an *underdetermination problem*: the underdetermination of concepts by concerns.

In sum, the general idea that our concepts should be answerable to our concerns leaves a great deal of work to be done before it can itself be put to work. We need to identify the relevant concerns for given concepts, and something needs to mediate between those concerns and the concepts they call for so as to give the concepts sufficiently determinate contours. What is more, all this needs to be achieved in a way that allows for generalizability and critical distance.

To make headway and identify the missing link connecting concepts and concerns, I propose to substitute the question: 'What do we aim to do in using this concept?' with a different question: 'What is the concept we *need* here?'

This points us towards a *needs-based* approach to concept appraisal that can overcome

all four problems we considered: the concern identification problem, the generalizability problem, the critical distance problem, and the underdetermination problem. For it is our needs for certain concepts that constitute the missing link mediating between our concerns and our concepts, and it is those needs that our concepts are, in the first instance, answerable to—or so I shall argue in the next chapter.

CHAPTER SIX

Tailoring Concepts to Needs

6.1 Conceptual Needs

‘If I have one advantage’, Nietzsche wrote, ‘it is a keener vision for that hardest and trickiest form of *backward inference* … from every way of thinking and valuing to the commanding *need* behind it’ (2005a, ‘Antipodes’). Nietzsche was right to see needs behind concepts. It is our *conceptual needs*—our needs *for* certain concepts—that form the conduits between our concepts and concerns.¹ If we manage to gain a sense of what the needs behind our concepts are, we can appraise concepts according to how well they meet those needs.

But what does it mean for a concept to have ‘a need behind it’? Making sense of conceptual needs does not require us to assume that concepts are the sort of thing one *simply* needs the way human beings simply need water or sleep. We must distinguish between *inner* needs and *instrumental* needs.²

¹ The term ‘conceptual need’ is used in a similar sense by Kappel (2010, 72). Earlier uses of the term in philosophy notably include Moravcsik (1976, 337) and Buchler (1955, 118). Aside from its prominent role in Nietzsche, a related notion might be thought to be implicit in the work of John Dewey, who models his theory of inquiry on the way biological organisms adjust to situations of ‘need’, ‘tension’, or ‘disturbed equilibration’ (1938, 6–7, 27–29); see Henne (2022, 10).

² This terminology combines Williams’s (2011, 51) contrast between ‘inner’ and ‘technological’ needs with Wiggins’s (2002, §6) contrast between ‘categorical’ and ‘instrumental’ needs. While Wiggins aims to elucidate how categorical needs make specially demanding claims on us, my focus lies on the class of instrumental needs characterized by the fact that what is needed, to whatever end, is a concept.

Inner needs can be physiological or psychological, but they are needs one has categorically, just in virtue of the kind of creature one is or has become (one can acquire inner needs). The physiological needs for air, water, and sleep, or the psychological needs for love and social interaction, are examples of inner needs. Ascriptions of inner needs do not invite the question of what one needs something *for*—one simply needs it.³

Instrumental needs, by contrast, involve a weaker, but far more pervasive form of needing: needing something as a means to the realization of some ulterior end, or as a remedy to an inconvenience. In this sense, one might need an umbrella when caught in the rain; one might need glasses to drive; or one might need patience when dealing with someone. Unlike ascriptions of inner needs, ascriptions of instrumental needs make it appropriate to ask what one needs something *for* ('I need a top hat.' —'What for?' —'For the play'). Instrumental needs are thus hypothetical rather than categorical: one has them *if* one is to realize some further end.

The needs for certain concepts, i.e. conceptual needs, are needs of the latter, instrumental kind. Concepts are needed, when they are, as a matter of technological necessity. They are needed instrumentally to perform some task, or to satisfy some prior need which may or may not be an inner need (not all instrumental needs have to be the products of inner needs—they can arise out of aims, projects, and other kinds of human concern).⁴ To say that we need a certain concept, or that we need a concept to take a certain form, is therefore elliptical: we need it *if* we are to perform some task that cannot very well be performed without it. This analysis leaves the notion of a conceptual need quite broad, because the underlying notion of a concern is something of a catch-all term encompassing any kind of motivation, desire, or commitment to a value or project whose realization makes demands on those who pursue it. But this is as it should be, since the range of ways in which concepts can serve people's needs is itself broad.

³ This may reflect no more than the fact that these needs can be attributed without further explanation, because they are extremely widely shared human needs whose intelligibility we take for granted. To add that one needs water or air *to survive*, or that one needs love or social interaction *to avoid serious psychological harm*, is not so much incoherent as uninformative. It may, however, help demystify inner needs; Wiggins (2002, 10), for example, suggests that inner needs can be demystifyingly understood as a special class of instrumental needs, namely those one has if one is to avoid serious harm.

⁴ Wiggins (2002, 10) suggests that inner needs can themselves be understood as a special class of instrumental needs, namely those one has if one is to avoid serious harm.

To a first approximation, the relevant sense of ‘conceptual need’ can be set out in terms of the following equivalence:

A set of concept-users U has a conceptual need for concept F

if and only if

Concept F is needful for U

if and only if

The use of concept F is conducive to meeting a concern C that U identifies with

It may be thought that this is too weak, however, because needing implies necessity or indispensability. Thus, Harry Frankfurt writes: ‘Nothing is needed except for the sake of an end for which it is indispensable’ (1984, 2).

It therefore seems that some further clause must be added, to the effect that U could not meet concern C without F ; or that it is necessary, if U is to meet concern C , that U use concept F . The equivalence then reads as follows:

A set of concept-users U has a conceptual need for concept F

if and only if

Concept F is needful for U

if and only if

The use of concept F is conducive to meeting a concern C that U identifies with and that, given U ’s capacities, U could not meet without F under the circumstances

if and only if

It is necessary, things being what they are at time t , that if U is to meet concern C at time t'' , U use concept F at time t'

The fact that conceptual needs are a function of our concerns affects the character of the necessity involved, however. Where inner needs are concerned, we are in the grip of ineluctable necessity: we simply need something. In the case of instrumental needs such as our needs for concepts, by contrast, we need something only insofar as it is indispensable to meeting a concern. This has three implications: the need will only be as important as the concern that gives rise to it; the need can be escaped if the concern can be abandoned (which is not an empty condition; there are things we cannot help but be concerned with); and what is indispensable is not necessarily a certain concept in

particular, but could be a more or less narrow class of concepts that are equally suited to meeting the concern.

For the purposes of instrumental need attribution, moreover, the evaluation of claims of the form ‘It is necessary, if x is to happen, that y happen’ does not, even in principle, require one to consider *all* possible future scenarios from the time and context of assertion onwards in order to verify that, in every scenario in which x happened, y happened also. The notion of an instrumental need displays a context-sensitivity that strongly narrows the range of possible future scenarios in which the object needed must figure for the need attribution to be true.⁵

This comes out already in Frankfurt’s own example of a man ‘who feels like completing a crossword puzzle and who is unable to do so without looking things up’ (1984, 2). Given his concern to complete the puzzle, that man really does need a dictionary, Frankfurt affirms, because the dictionary is indispensable for him to complete the puzzle.⁶ But this does not entail that, necessarily, in all possible future scenarios in which the puzzle has been completed, it has been completed using the dictionary. It *could* conceivably be completed without dictionary—just not by this man. His need is thus sensitive to his abilities and limitations.

Moreover, even this man could complete the puzzle using some other means—perhaps an unexpected visitor happens to drop by, who, when desperately urged to help, turns out to be surprisingly knowledgeable. In that case, our man could reasonably exclaim: ‘You are just what I need!’ That would not falsify Frankfurt’s observation that the man needed the dictionary before the unexpected visitor showed up. It is merely that this earlier assessment of his needs was sensitive also to what means were realistically available to him under the circumstances. That hapless visitor, being unexpected, did not yet figure in the assessment.

In addition, need attributions are informed not just by a sense of what one can do and what is available to one, but also by a sense of what constitutes an acceptable means of satisfying one’s concern and what is beyond the pale. Our man *could* complete the puzzle by calling up his sister whom he knows to have completed it already. But that would be cheating. Similarly, the child that tells its parents that it needs money to buy a new

⁵ Wiggins (2002, 12) also registers this feature of need attribution.

⁶ See Frankfurt (1984, 2, 7–8).

calculator has not overlooked the possibility of stealing one. What one needs to satisfy a concern are the most promising means that are left over once the unrealistic and unacceptable options have been ruled out. Consequently, the clauses in the aforementioned equivalence concerning what one ‘could not do without’ the concept, or concerning the ‘necessity’ of using some concept to meet some concern, must be read narrowly. They are focused on scenarios that are realistic and acceptable given how things present themselves in a certain context at the time of the need attribution.

While this context-sensitivity makes instrumental needs in general highly variable, *conceptual* needs are much less variable due to the character of what they are needs for. Though they are a species of instrumental need, they are needs for concepts; and concepts are not single-use instruments of concern satisfaction. We adopt concepts not the way we seize on the nearest sharpish object suitable to be abused as a letter opener, but rather the way we settle on a policy: we settle on a stable, rule-governed pattern that we stick to across a range of applications.⁷

Just because of its stability and rigidity, concept use will, like adherence to a policy, involve trade-offs, drawbacks, and pitfalls. But its settled character is also what makes a concept shareable, teachable, and reliable. It enables a concept to be efficiently deployed, taken for granted, and relegated to the background as we focus on the object-level considerations it brings to our attention.

6.2 Needfulness Conditions

Thinking of conceptual needs as instrumental needs entails that they are not simply given in virtue of the kinds of creatures we are. Conceptual needs must themselves be understood as products of the characteristics of concept-users and their situation. One does not fully grasp an instrumental need unless one grasps what engenders it.

This indicates the crucial idea that a given concept is only worth using if certain extraconceptual presuppositions are fulfilled: most obviously, concept-users must pursue certain *concerns*, and the concept must be suitable to meeting these concerns. Notice that the catch-all term ‘concerns’ allows us to widen the scope of relevant input to

⁷ On the practical exigencies that underlie this feature of concepts where moral concepts are concerned, see Sinclair (2021, 98–100).

include more than just the goals or conscious aims with a view to which concept-users employ concepts.

As highlighted by the problem of the underdetermination of concepts by concerns, however, the specification of conceptual needs requires more than just concerns. It also requires factoring in concept-users' capacities and limitations as well as their circumstances. Only then can we meaningfully speak of concept-users as having conceptual needs. What our conceptual needs are is not simply a function of our concerns, but also depends notably on our capacities and on how the world is, and more particularly on what circumstances demand of one who pursues that kind of concern with those capacities.

By way of illustration, consider a very simple-minded creature.⁸ It has only one concern, which is to eat. It has the capacity to seize what food it can ingest when it immediately presents itself, but is otherwise immobile. And its circumstances are such that food is not constantly abundant, but only occasionally presents itself. This triad of concerns, capacities, and circumstances already suffices to create an instrumental need in our simple-minded creature: it needs a bundle of dispositions that are reliably and differentially responsive to the presence of the relevant stimuli, enabling it to actualize its capacity to seize food when and only when it passes by.⁹

Is that already a need for anything like the concept *food*? Hardly. Our creature is not yet a concept-user, but merely a reliable differential responder—comparable perhaps to a carnivorous plant like the Venus flytrap. And what it responds to is not some particular aspect of a conceptually articulated experience of something beyond itself, but the situation as a whole: the conjunction of its own concern and capacities with certain external circumstances. There is no practical pressure here to impose a conceptual articulation on experience, and distinguish between its own capacities, its own concerns, and things out there that will or will not meet its concerns given its capacities.

But give the creature the capacity to direct its movements and roam around in search of food, and you increase the practical pressure on this 'primitive holism' to start to

⁸ I am indebted here to descriptions of similarly simple-minded creatures in Bennett (1976), Lloyd (1989), and especially to Craig (1990, 82–84).

⁹ I take the notion of a reliable responsive differential disposition from Brandom (2002a, 350), who develops it on the basis of Sellars's account of observational knowledge; see also Brandom (2015b, 101).

'fragment', as E. J. Craig (1990, 83) puts it. If, instead of roaming around at random, such a mobile creature became able to distinguish between 'food, here, now' and 'food, here, not long ago', or 'food, here, soon', it could dramatically improve its odds of satisfying its concern to eat. But it would also come under strong practical pressure to become sensitive to the difference between circumstances that its own capacities and limitations equip it to handle and circumstances it is not equipped to handle: that difference can be further differentiated along many dimensions, but already the failure to draw the basic distinction between 'food, over there, that I can get to' and 'food, over there, that I cannot get to' would soon cost it its life.

Already this simple example brings out how concerns combine with certain capacities and circumstances to generate certain instrumental needs, which, in social and language-using creatures like us, would take the form of *conceptual* needs. The concerns of concept-users must combine with their capacities and circumstances to render a concept, or some broader class of concepts of which it is an instance, *needful*. What it renders needful, in particular, is that which will allow one to satisfy the relevant concern, given one's capacities and circumstances.

Before the issue of whether a concept is more or less *needful* than another can even arise, however, any concept first needs to pass the minimal threshold of being so much as *useable* at all, i.e. fit to be used by real human concept-users. Concepts do not apply themselves; nor are they applied by all-seeing disembodied intelligences, capable of telling, immediately and definitely, what any imaginable concept is applicable to. Approaching concepts as if from the perspective of a disembodied and omniscient intelligence that can instantly compute the extension of any given concept for any world or world-time pair, as is typically done in formal semantics,¹⁰ is helpful in trying to render the semantics of natural languages mathematically tractable, or in showing how the meanings of complex expressions can be computed from the meanings of simpler expressions. But to evaluate how concepts actually fare in the service of real concept-users, we have to relate the concepts to concretely situated agents with certain concerns and capacities, encountering a particular world from a particular perspective that

¹⁰ David Chalmers, for example, writes: 'We can say that a subject grasps an intension when the subject is in a position to *evaluate* that intension: that is, when sufficient reasoning will allow the subject to determine the value of the intension at any world' (2002, 148).

imposes serious epistemic limitations on them—most basically, the limitations of only ever being at one place at a time, and of perceiving the world through a physiological constitution that makes it hard to apply concepts at certain scales and through certain sensory modalities.

From the perspective of such concept-users, one of the most fundamental challenges concept-users face is precisely the challenge that is elided by the perspective of the omniscient computer of extensions across all possible worlds, namely the challenge of how to *tell* whether a given concept applies to something.¹¹ This already makes demands on concept-users: the correct application of concepts typically demands some *investigative effort* to get into a position from which to tell whether a concept applies, be it in the form of reasoning (e.g. does the concept *prime number* apply to 514229?) or in the form of changing one's vantage point in time or space (e.g. does the concept *ripe* apply to yonder fruit?). But it also makes demands on concepts themselves—demands we only see once we consider the challenges involved in *operationalizing* concepts for use by practically situated concept-users.

For concepts to be capable even in principle of rewarding investigative effort, they need to be *recognizably applicable*: they need to be such as to hand their users some means of *tracking* what the concept picks out. Some concepts, such as the concept *blue light*, make this easy: the properties picked out by the concept coincide with those by which users of the concept typically identify its extension—in which case we might say that the properties guiding use of the concept of *F* are not just *indicative*, but *constitutive* of *F*. This is true of all concepts that resist the distinction between *seeming* to be *F* and *being F*. (Even if you wear blue-tinted glasses, the light you see does not just *seem* to be blue; the light going from the glasses into your eyes really *is* blue light if anything is.)

But many other concepts, in order to be world-guided in their application at all, need to rely on discernible properties that are indicative of *F* *without* being constitutive of *F*: properties that can be tracked in order thereby to track *F*. A property indicative of *F* could be any property *G* such that things that are *G* tend to be *F*. Of course, the concept of *F* is

¹¹ For a related point, see Brigandt (2013, 76), who observes that a subject may lack the ability to evaluate the intension of a concept in a given world for lack of other concepts: geneticists in the 1930s had the concept *gene*, but were unable to determine its exact extension because they lacked the necessary concepts of molecular biology.

correctly applied only to things that really are *F*, but in order to tell which these are, we need to rely on indicators which give us *reasons* to judge that a given thing is *F* (—‘What makes you think this is a Rembrandt?’ —‘Just look at the use of directional light and the carefully detailed facial expressions’).

To epistemically limited concept-users, these more accessible or more recognizable proxies for what is to be picked out offer indispensable practical assistance. We must only be careful not to equate indicative with constitutive properties: we tell the time by measuring the movement of clock hands, but this does not mean that time *is* the movement of clock hands; we identify Mars in the night sky by looking for a red planet, but this does not mean that Mars *is* a red planet; we attribute mental states to others on the basis of observable behaviour, but, *pace* Skinnerian behaviourists, this does not mean that mental states *are* observable behaviour; we discern causation notably by tracking constant conjunction, but, *pace* the constant conjunction theory of causation, this does not mean that causation *is* constant conjunction.¹²

This suggests a very basic desideratum for concepts: the concepts we need should track their extensions via *reliable* indicators. That is to say, a good concept licenses reliable inferences from the presence or absence of certain indicator properties to the applicability or inapplicability of the concept.

This conceptual reliability can in turn be analysed into two aspects: *conceptual sensitivity* and *conceptual specificity*. Let conceptual sensitivity be the degree to which the indicator properties that a concept tracks its extension by form a reliable way of identifying that which *does* fall under its extension (if x is *F*, how much of the time does a user of concept *F* recognize that x is *F*?). Low conceptual sensitivity will then favour ‘false negatives’: the concept-user will be prone to judge that x is *not-F* when, in fact, x is *F*, thereby missing many instances in which the concept applies.

Conceptual specificity, on the other hand, can be defined as the degree to which the indicator properties are reliable ways of excluding what does *not* fall under the concept (if x is *not-F*, how much of the time does a user of concept *F* recognize that x is *not-F*?).

¹² On that last point, see Harré (1964, 359), who charges Hume with making this mistake. While Hume certainly took the observation of constant conjunction to be an important *source* of our concept of causation, however, it is less clear that he meant to proffer an *analysis* of the concept in terms of constant conjunction.

Low conceptual specificity will favour ‘false positives’: the concept-user will be prone to judge that x is F when x is $\text{not-}F$, treating the concept as applicable even in instances in which it only *seems* to be applicable without really *being* applicable.

Analysing reliability into sensitivity and specificity allows us to see the differences in value between concepts that are less than fully reliably to the same degree but along different dimensions. Consider the concept *fraud* as employed by administrators in assessing claims to unemployment benefits, for example. If those administrators’ capacities to investigate and access information about each individual claimant are limited, their application of the concept will need to be guided by indicator properties that are less than fully reliable. But there are two ways in which that could be so. The concept could be somewhat unreliable by being applied *too narrowly*, i.e. generating false negatives due to a lack of sensitivity, or by being applied *too broadly*, i.e. generating false positives due to a lack of specificity.

This distinction between two species of unreliability clearly matters to the appraisal of the concept: if one concept of fraud is applied too narrowly to some degree and another concept of fraud is applied too broadly to exactly the same degree, these two concepts of fraud will be equally unreliable but differ in the kinds of consequences their unreliability has. If the concept is applied too narrowly, the state loses money over fraudulent benefits claims. If the concept is applied too broadly, the state leaves people to starve. That this is a significant difference is agreed even between those who disagree about which of these outcomes is preferable.

Given this analysis of conceptual reliability into conceptual sensitivity and specificity, it might be tempting to conclude that we always need our concepts to be *maximally* sensitive and specific, leaving no room for discrepancies to open up between the indicator properties and what they indicate. This is what might be called the ideal of the *fail-safe* concept: a concept such that whenever one genuinely has reason to think it applies, it really does apply, and it never applies without there being reason to think it applies. Admittedly, a human mind stocked solely with such fail-safe concepts might still introduce mistakes into the process of deploying those concepts, for instance by erroneously taking itself to have reason to apply a concept. But such a failure would be the result of human error, not of conceptual unreliability.

If we take seriously the idea of approaching concept appraisal from the vantage point

of practically situated concept-users trying to make the most of their limited capacities, however, this ideal of the fail-safe concept is clearly misguided. We *need* there to be potential gaps between indicators and what they indicate, because the elusive world we inhabit would largely remain silent before a mind stocked solely with fail-safe concepts, leaving us poorly equipped to make much useful sense of the world around us.

Thinking about conceptual needs from the perspective of practically situated concept-users will thus help us narrow down what concepts we need precisely by taking into account the various contingent limitations and constraints imposed on us by our capacities and circumstances. We need certain concepts to help us satisfy our concerns precisely because we face situations that make it in various ways challenging to satisfy them, and often require trade-offs.

Approaching concept appraisal from such a needs-based perspective will therefore not always favour the concepts that are the most reliable. As enticing as the idea of a concept tracking infallible indicators sounds, we need to strike a balance between reliability and usability. A concept can more than make up for its relative unreliability by being eminently usable, tracking properties that are readily accessible. Reliability is a good, but it is not the only good. Moreover, a concept can compensate for its unreliability through its *flexibility*: the degree to which a concept can accommodate the variability of indicator properties. If the properties indicating a concept's applicability vary widely, but most of these indicator properties are not strictly necessary conditions for something to fall under the concept, we are better off using a concept that 'rides such variability', in Adrian Moore's (1993, 286) phrase, tracking the range of properties that typically, if defeasibly, indicate that something falls under a concept.

We might call all these contingent, extraconceptual conditions that are jointly sufficient to render a particular concept needful the concept's *needfulness conditions*. To achieve a better analytic and mnemonic grip on needfulness conditions, we can think of them as always combining three aspects:

- (i) the *concerns* of concept-users, i.e. what they care about (because they categorically need, want, or value it);
- (ii) the *capacities* of concept-users, in the inclusive sense encompassing their corresponding limitations; in particular, the limited physiological, perceptual, and cognitive capacities that they have upstream of adopting the concept at issue, but

- also the cultural and technological resources they can draw on;
- (iii) the *circumstances* in which concept-users seek to satisfy these concerns with the limited capacities they have, i.e. the natural and social environment in which the concept is to be deployed; this might include their geographical situation, their material and technological affordances, their social structures and institutions, and their position in society.

Conceptual needs thus arise out of the interaction between our concerns, capacities, and circumstances: what demands our concepts should make on our thought depends on what demands the world makes on us given the demands that we make on the world.

The implication is that concepts come with certain extraconceptual presuppositions that have to be realized for concepts to be needful. Only when these needfulness conditions are fulfilled is there a point to using a given concept, because it is these conditions that imbue the concept with *pointfulness*—the concept's capacity to help us to live by making a useful difference to our lives. The concept of weight, for example, is one we instrumentally need for all kinds of ulterior concerns, and yet it would be pointless if the laws of gravity were such that objects on earth randomly and radically changed weight all the time.¹³ Conversely, the concept of health would be pointless for us if our physiological condition never changed. Or take the concept of intention, which, as G. E. M. Anscombe observed, would be pointless if we took absolutely no interest in each other's reasons for action.¹⁴ And as Wilde's refusal to so much as enter into whether a certain story was blasphemous might be taken to imply, the concept of blasphemy is pointless if there is no deity out there to offend.¹⁵

While this explicates the pointfulness of a concept in terms of its needfulness—what gives point to a concept is the fact that it meets an instrumental need—needfulness is not the same as pointfulness: the property of being needed is notoriously insufficient to bring its bearer into existence. As Jeremy Bentham observed, 'a reason for wishing that a certain right were established, is not that right—want is not supply—hunger is not bread' (1843, 501). A concept can thus be needful long before it comes into use. It can only become

¹³ Wittgenstein (2009, §142) offers a similar example.

¹⁴ See Anscombe (1957, §21).

¹⁵ See Hyde (1973, 107).

pointful once it is in use, however, and it will be pointful if and as long as it is instrumentally *needful*.

The presuppositions that give point to a concept thus revolve around what users of the concept and the world in which it is deployed are like. But these presuppositions remain extraconceptual: they do not have to figure in the *content* of the concept; nor do they have to be *constitutive* of the concept. It is merely that living by a certain concept only has a point, i.e. fills a need, against the backdrop of certain facts: the facts engendering an instrumental need for the concept. This holds quite generally, as Wittgenstein emphasizes: ‘if anyone believes that certain concepts are absolutely the correct ones, and that having different ones would mean not realizing something that we realize’, he writes, ‘let him imagine certain very general facts of nature to be different from what we are used to, and the formation of concepts different from the usual ones will become intelligible to him’ (2009, II, §366).

6.3 The Expressive Character of Concepts

Even on this account, however, a concept is not inherently needful, nor does it inherently fill a need: it only does so if hooked up to the right concept-users in the right circumstances. So how can one work back from a concept to the conditions that might render it needful and therefore pointful? This is the form that the concern identification problem takes for the needs-based approach.

It is my contention that what can guide us from concepts to the relevant concerns is the *expressive character* of concepts: a concept can *express* something of the conditions in which it *would* be worth using, even when these conditions are not given.

An emphasis on expressive character is perhaps most at home in the appraisal of works of art; but this does not mean, as Andrew Huddleston has stressed, that the grounds of appraisal guided by expressive character must themselves be aesthetic grounds.¹⁶ To berate a concept for being *cloying* would be to criticize it on aesthetic grounds. But we need not play Walter Pater at the museum of concepts. We can export a sensitivity to

¹⁶ See Huddleston (2019, 158). Though the account of expressive character I develop here is ultimately quite different from his, I would not have developed it without his work highlighting the expressive dimension of Nietzsche’s critique of values.

expressive character out of the aesthetic sphere without tying ourselves to aestheticism about ethics and politics.

What lends a concept its expressive character? Could it be a concept's *content*? Clearly, a concept's content contributes to determining its expressive character. But, in the sense at issue here, what a concept expresses is not to be identified with its content. On the contrary: what makes concepts such insidious ideological tools is precisely that they can express objectionable concerns *without* making any reference to these concerns at the level of their content. They would not be effective ideological instruments otherwise. In E. P. Thompson's example: 'If the law is evidently partial and unjust, then it will mask nothing, legitimise nothing, contribute nothing to any class's hegemony. The essential precondition for the effectiveness of law, in its function as ideology, is that it shall display an independence from gross manipulation and shall seem to be just' (1975, 263). Conversely, the most exalted concerns of liberal democracy might find their most important expression in the dry jargon and sterile concepts of bureaucrats and legal clerks. Part of the ingenuity of such arrangements is that this does *not* depend on those concerns figuring in the thinking of the bureaucrats and clerks in question. A concept's content and what it expresses can radically come apart. So how can we make sense of this elusive dimension of concepts?

One possibility would be to identify the expressive character of a concept with the *attitudes* that concept-users express through *tokenings* of the concept on particular occasions. This would allow conceptual content and expressive character to come apart: the thick aesthetic ideals of a tailor who judges the cut of a suit, for example, might only ever find expression in tokenings of very thin terms ('Just right!');¹⁷ likewise, the moral ideals of a writer might be forcefully—indeed, *more* forcefully—expressed in a soberly descriptive account of how someone pulls a horse out of deep snow.¹⁸ But this analysis would not entail that the concept *right* itself inherently expressed sartorial ideals, or the concept *horse* moral ones. Rather, the concerns expressed by those concepts would simply be the concerns of *particular concept-users* as expressed in individual instances of concept use. For an account of the expressive character of *concepts*, we therefore need to look elsewhere.

¹⁷ The example is Wittgenstein's (1966, 5–9).

¹⁸ See Diamond (2018, 225–9).

On the account I propose, the key to the expressive character of concepts is to see that even when a concept fails to fill an instrumental need, because it is not currently hooked up to the right concept-users in the right circumstances, it nonetheless *expresses* the conditions that *would* render it pointful. One merely has to ask: who would have need of such a concept? What combination of concerns, capacities, and circumstances would give one reason to think in these terms?

By tentatively assuming that a concept answers to an instrumental need, we can work our way back to conditions from which such a concept would derive a point. There is no guarantee that we will not come up empty. But this assumption of instrumentality functions like an interpretative ‘principle of charity’ for needs-based concept appraisal. It is the hermeneutic lever that gets needs-based appraisal off the ground, and allows us, when successful, to work our way to a picture of the conditions from which such a concept would derive its point. These are the conditions that the concept might be said to *express*, even when these conditions are not presently given. The conditions a concept expresses in this way are, in the first instance, its *pointfulness conditions*, but since its pointfulness in turn presupposes its needfulness, it thereby also expresses its *needfulness conditions*.

Strictly speaking, what a concept expresses in this way is always the entire *set* of conditions that would jointly engender a need for it and imbue it with a point: not only (i) some human concern, but also (ii) the limited capacities of concept-users with which they pursue that concern, and (iii) the circumstances in which they do so; for only when joined together into a set do those three kinds of conditions give those concept-users reason to pursue that concern using that kind of concept.

But one can—and Nietzsche, for one, often does¹⁹—take a concept to *metonymically express* one element of such a set by expressing the whole set: a concept tailored to satisfy a concern for revenge in the hands of the weak under circumstances of oppression by the strong, for example, might metonymically be said simply to express a concern for revenge; or to be a sign of weakness; or to speak of circumstances of oppression. When Nietzsche asks, in Third Treatise of *On the Genealogy of Morality*, ‘What do ascetic ideals

¹⁹ See Queloz (2023) for a detailed account of Nietzsche’s expressivist critique of concepts along the lines presented here, and see Huddleston (2019) for a different reading that also emphasizes the expressive dimension.

mean?', the answer he eventually gives suggests that what he is after is what they *express*: '[a] hatred of the human, and even more of the animal, and more still of the material, [a] horror of the senses, of reason itself, [a] fear of happiness and beauty, [a] longing to get away from all appearance, change, becoming, death, wishing, from longing itself' (1998, III §28). Clearly, all these attitudes and concerns are not literally part of the *content* of ascetic ideals. Rather, the thought is that if we let ourselves be guided by the question of *who* would have need of ideals such as these, we are led to the conclusion that it is *for people with this particular bundle of concerns* that living by such ideals would be most pointful.

We can thus understand the expressive character of a concept in terms of how the concept expresses certain presuppositions, namely its pointfulness conditions—the conditions that would render the concept instrumentally needful and thereby give point to its use.

To be sure, moving from a concept to its pointfulness conditions requires a genuine backward *inference* that requires delicate interpretation and judgement, and there may not be a uniquely right way to do it. Some concepts plausibly serve several points at once. Certainly, a concept's pointfulness conditions cannot simply be read off the concept algorithmically, as a barcode scanner might read the number of a product. Discerning a concept's expressive character remains a matter of interpretation.

But that interpretation is nonetheless constrained by what genuinely makes sense and what remains obstinately unintelligible, what holds up and what falls flat. If future generations subsisting exclusively on a diet of sustainably produced vitamin pills were to dig up one of our can openers, they might reasonably hypothesize—even if they no longer store food in cans—that this object expresses an instrumental need to open cans, and thereby expresses conditions of life that render it instrumentally needful to store food in cans; the rival hypothesis that the can opener expresses a need to make music will simply not hold up in light of the object's intrinsic features and the kinds of sounds humans tend to regard as musical. More generally, there are claims about what a concept expresses that no truthful inquiry into the concept's inherent properties and their likely impact on human affairs will bear out. This is what sets off ascriptions of expressive character to a concept from projections of subjective associations onto a blot of ink.

What it takes to substantiate a claim about expressive character, in particular, is to

make plausible that the concept in question is *apt* to serve the concerns of certain types of concept-users under certain types of circumstances, and that this aptness is primarily due, not to an improbable alignment of the stars, but to features of the concept itself. To be able to make sense of a concept as expressing a certain concern, we have to be able to envisage what would engender an instrumental need for something very like this concept. The envisaged situation need not be, nor ever have been, actual. A concept's aptness for serving certain concerns might in principle be revealed to be latent in a concept merely by considering counterfactual situations.

On this account, then, to consider a concept's expressive character is to think in *modal* terms about what kinds of concerns it would be most apt to help to meet, and not just to look at what concerns it *actually* helps to meet. To say that a concept expresses a concern for revenge, or a longing for another world, or a yearning for life to be ultimately fair, is to say that the concept is recognizably tailored to *serving* those concerns, which is to say that, in the right hands and under propitious circumstances, its use would tend to have effects conducive to the satisfaction of those concerns.

We can thus distinguish two different relations in which concepts can stand to concerns, on the present account. A concept can *express* a concern or *serve* a concern. In order to identify whether a concept serves any of our concerns, I have suggested, we can ask what concerns it expresses.

In the conceptually basic case, a concept also serves the concerns it expresses. But the two relations can come apart: a concept might *express* a concern *without serving* it, because the conditions necessary to serving it are not given; and it might *serve* a concern *without expressing* it, because its effects only fortuitously but unsystematically and unreliably satisfy that concern through some fragile alignment of circumstances.

What concern a concept expresses is a function of the robustness of its ability to serve the concern across variation in the relevant contingencies. A concept need not *invariably* serve a concern in order to express it. But it does need to serve it *non-accidentally*. There has to be something about the concept itself that makes it *apt* to serve that concern.

One way of spelling out this idea is again in modal terms, which we can do particularly

vividly in the idiom of ‘possible worlds’:²⁰ there has to be, not just a handful of scattered because vastly different possible worlds in which it happens to serve that concern, but a reasonably large cluster of neighbouring possible worlds in which it systematically, if not invariably, serves that concern, even if our actual world now lies outside that cluster.

By way of illustration, consider the legal concept of a basic rights infringement, which, when demonstrably satisfied, empowers individuals to trigger a process of judicial review aiming to determine whether their basic rights have in fact been violated by state action or omission. This concept plausibly evinces a liberal concern to protect individual liberties against the powers of the state. In the first instance, therefore, the concept might be thought to express a liberal society’s instrumental need to give individuals legal means to push back against the state’s curtailment of their most fundamental liberties.²¹ Still, by some unlikely conjunction of circumstances, the concept might end up playing into the hands of the *illiberal*-minded at a certain juncture. Does this mean that the concept also expresses their concerns? No, because this serviceability is of the accidental kind ‘which alters when it alteration finds’, as Shakespeare has it: it is highly counterfactually fragile, breaking down in most nearby possible worlds. The concept’s aptness to serve the concerns of the *liberal*-minded, by contrast, is far more counterfactually robust—that is what lends force to the claim about the concept’s expressive character, regardless of whether the liberal concern is actually being satisfied as things currently stand.

This account of the expressive character of concepts might be thought to resemble accounts of ‘expressive meaning’ in non-truth-conditional semantics, which claim that the expressive meaning of words such as ‘Ouch!’ or ‘cur’ can be captured by identifying the conditions for the ‘felicitous use’ of those words: ‘Ouch!’ is felicitously used just in case the speaker experiences pain; ‘dog’ and ‘cur’ have the same truth conditions, but ‘cur’ adds the ‘felicitous use’ condition that the speaker have a negative attitude towards the

²⁰ The use of this idiom need not entail unpalatable ontological commitments to possible worlds and their equally uninviting implications for modal epistemology. Talk of possible worlds can be given a deflationary gloss as an especially perspicuous and precise way of making explicit what ordinary modal talk expresses anyway. Like ordinary modal talk itself, talk of possible worlds can be demystified by an account on which it ‘serves the function not of tracking features of additional worlds and reporting on their features, but rather of adding … expressive power to our language’ (Thomasson 2020a, 123); see also Brandom (2008).

²¹ See Cueni (manuscript-b) for a detailed argument to that effect.

referent.²²

But the account of expressive character offered here differs from these semantic accounts in several respects. First, the present account is not about the meaning of words, but about the pointfulness of concepts. Second, the expressive relation it highlights between a concept and its presuppositions is a functional rather than a semantic relation: the concept can only serve its point if the presuppositions are realized, but this functional relation need not be part of what a competent user would have to grasp about the concept in order to count as competent (whereas someone who did not understand that using the word ‘cur’ felicitously presupposed a negative attitude on the speaker’s part would fail to grasp the full meaning of the word). And third, needs-based concept appraisal aims, in the first instance, to intuit what makes the entire practice of living by a certain concept pointful, not what makes individual instances of concept use pointful.

At the same time, it is also true that by grasping what makes a practice of concept use pointful in general, one acquires some sense of when it would be pointful in particular. Consider the cyclists’ practice of holding out their right hand before taking a right turn: in virtue of my understanding of what makes the practice pointful in general, I become able to recognize its pointlessness on particular occasions.

Thus, what a concept expresses of the conditions that would render it pointful is a valuable epistemic guide to the concerns in relation to which it should be appraised, even if no-one presently has the concerns that would be best served by such a concept. By coming to understand that a certain concept is the kind of concept that people driven by certain concerns would have a need for, one is given some reason—however defeasible—to think that the effects of living by that concept, even when *lacking* such concerns, are unlikely to be conducive to the satisfaction of different concerns, and more likely to further, *even in their absence*, the concerns of those who would have need of such concepts. That is why Nietzsche was alarmed about the prospect of Christian values remaining in use even after the ascetic priests who promulgated them had lost their influence. Concepts can subsist outside of their needfulness conditions, and thereby continue to systematically have effects that we might not want to see systematically realized. Examining a concept’s expressive character is thus a particularly instructive way

²² See Gutzmann (2013) for a survey of varieties of expressive, non-truth-conditional meaning. A closely related approach is ‘success semantics’ (Blackburn 2005).

of assessing a concept's causal profile, because it enables us to contemplate not just the effects a concept actually has, or has actually had, but the effects it is likely to have going forward, just in virtue of the kind of concept it is. By asking who would be best served by a concept, one learns something about whether it is likely to serve us.

6.4 Need Matrices

On this account, then, it is to the welter of our concerns, capacities, and circumstances we must look to appraise our concepts, and we receive some guidance in this from the expressive character of concepts. But this still leaves us with a formidable epistemic challenge. For, on this account, what needs our concepts are answerable to is, in the end, a function of all the concerns we identify with, all the capacities we have or lack, and all the circumstances we face. Philosophers of a certain temperament will welcome the implication that we have to engage with the messy complexities of social reality in order to determine how our concepts relate to our concerns. But this thicket of need-engendering conditions threatens to be intractably dense.

In some instances, we may hope to overcome this epistemic challenge by clearly delimiting a set of *use cases* for a concept. This might notably be achieved by focusing on the concepts needed for particular roles within particular institutions. In *When the State Meets the Street* (2017), for example, Bernardo Zacka draws on his ethnographic fieldwork to provide a comprehensive picture of the conceptual needs of social workers at the frontline of a social welfare agency.²³ This clearly delimits a set of relevant concept-users, concerns, capacities, and circumstances. These social workers are concerned to organize their work effectively and efficiently, be consistent in how they handle cases, exchange advice, and teach new workers. As a result, Zacka argues, they need concepts that define an informal taxonomy of types of clients and responses they call for. He concludes that a conceptual framework allowing them to ascertain and communicate whether clients have a *situation*, *issues*, or an *attitude* goes a long way towards meeting their conceptual needs: clients with a *situation* require prompt attention; clients with *issues* face serious personal difficulties that provide some excuse for inappropriate behaviour and demand forbearance; clients with an *attitude*, by contrast, are prone to

²³ See Zacka (2017, 152–99).

exhibit inappropriate behaviour without any excuse for it.²⁴

This informal taxonomy is a far cry from a full-blown normative theory that could motivate, justify, and harmonize all street-level operations of the state. But given their capacities and circumstances, these social workers do not have time to systematically reason their way back to first principles on each occasion in the manner of Dworkin's ideal judge Hercules.²⁵ A modest typological systematization through an informal taxonomy does more to meet their needs. It leaves them better able to respond to the particularities of individual cases, faster in deploying these more lightweight cognitive resources, more flexible in accommodating new types of cases, and quicker to revise their conceptual framework when appropriate.

In the kinds of cases that are perhaps most characteristic of philosophical reflection on concepts, however, the concepts at issue are often too widely applicable to be appraised on the basis of a well-delimited and suitably narrow range of use cases: the concepts *truth*, *knowledge*, *understanding*, *justice*, *equality*, *liberty*, or *voluntariness*, for example, are not the preserve of a particular agency. The concerns that these concepts make contact with across different situations are likely to be mindbogglingly multifarious. How can we render this seemingly intractable welter of concerns, capacities, and circumstances philosophically tractable?

I propose that we approach that messy complexity piecemeal, by building a series of incomplete but illuminating philosophical models of it, each crystallizing and holding up for philosophical scrutiny a minimal set of concerns, capacities, and circumstances sufficient to generate a conceptual need. This is broadly in line with the conception of philosophy as model-building advocated notably by L. A. Paul (2012) and Timothy Williamson (2017; 2018a, 130–40; 2018b; 2020, ch. 10). ‘Humans are a classic example of messy complex systems’, Williamson observes, which is why ‘one might expect a model-building strategy to be appropriate’ (2018a, 130). Different types of models, admitting of varying degrees of rigour, will be appropriate to different areas and approaches. But for a needs-based approach to concept appraisal, what we want is a model that isolates a set of needfulness conditions and highlights a resulting conceptual need that our conceptual repertoire should be responsive to. This may merely give us a

²⁴ See Zacka (2017, 163–64).

²⁵ See Dworkin (1986, 239).

partial picture of a concept's total needfulness conditions. But it gives other philosophers a well-delineated set of criticizable claims to focus on, scrutinize, and complement or improve on with models of their own.

For any given concept, we can try to build a model of its needfulness conditions by asking: who would have need of such a concept? What combination of concerns, capacities, and circumstances would give one reason to reason in these terms? By using as our hermeneutic lever the assumption that a concept answers to an instrumental need, we can perform a backwards inference to the conditions from which such a concept would derive a point. This assumption of instrumentality enables us, when successful, to construct a model of a concept's needfulness conditions.

I shall refer to this type of model as a *need matrix*: an array of concerns, capacities, circumstances jointly sufficient to generate a conceptual need. A need matrix isolates a particular set of needfulness conditions and highlights a conceptual need that this set of conditions suffices to engender. The term 'matrix' helpfully connotes three ideas: that we are dealing with a whole array of intersecting concerns, capacities, and circumstances; that they combine to generate something, namely a conceptual need; and that, like the printer's matrix, a need matrix can act as a mould for casts of thought.²⁶

In light of a need matrix, a concept will emerge as apt or inapt in particular respects, as the matrix conveys a sense of what kind of concept we need and why: what the concept needs to pick out, what inferential connections it needs to allow us to draw, and how it needs to channel motivation to result in the effects needed to realize concept-users' concerns. Using a task matrix, we can arrive at a fine-grained and case-specific sense of what concept best handles a particular combination of practical pressures. This allows us to substitute the question of what makes an inherently good concept, or a good concept *tout court*, for another, more determinate question: what makes a good concept for concept-users with those concerns and capacities in those circumstances?

A useful heuristic in constructing a need matrix is to start with a very generic matrix

²⁶ The etymology of the term 'matrix' traces back to the Latin *mater*, meaning 'mother' or 'womb', but the term is commonly used more metaphorically, to refer to the environment out of which concepts develop; to describe an organizational structure in which an array of lines of command or responsibility run through a single person (in allusion to matrices in mathematics); and as a name for the moulds used for casting printing type. All three metaphorical senses are relevant to the notion of a need matrix.

modelling some of the more general human concerns to which a concept answers before factoring in how this abstract schema has been *elaborated*, i.e. inflected, extended, and redirected, by the historical, social, and cultural forces that characterize our own more concrete circumstances. It is only through their elaboration by a particular sociohistorical situation that inchoate and indeterminate concerns are concretized and given a definite direction. If we want to appraise and improve a particular part of our conceptual repertoire, we accordingly have to arrive at a fairly nuanced understanding of the relevant concerns and of what they are now directed to. We can get started on this task by reconstructing a basic need matrix that plausibly captures something about why we first came to be concerned with anything like that concept to begin with.

In the case of the political concept of liberty, for example, the most basic need matrix out of which that concept grows might be thought to centre on the fact that human beings care about not being frustrated by other human beings in the realization of their desires (an idea I expand on in Chapter 9). But, of course, this basic concern for freedom takes us nowhere at all when, for instance, we want to assess the authority of Philip Pettit's and Quentin Skinner's efforts to move us from a concept of *liberty as non-interference* towards a concept of *liberty as non-domination*. To make the case for the latter concept using a needs-based approach would require one to show that this concept better meets our conceptual needs as expressed in the context of twenty-first century liberal democracies.

What this combination—of a basic task matrix and a certain sociohistorical elaboration of it—aims to do is to yield the kind of understanding that manages to hold on to two important thoughts at once: that many of the concepts philosophy grapples with are rooted in highly general human concerns that are very widely shared across human history; but that these concerns nonetheless find expression in our own time and place in highly specific and peculiar ways that are certainly not widely shared across human history. Evolutionary psychology and game theory are apt to emphasize the first thought, but tend to do so at the expense of the second; history and sociology are apt to emphasize the second thought, but tend to do so at the expense of the first. To get the full measure of a concept's value to us, we need to hold on to both thoughts at once.

The methodological constraint of having to specify a need matrix structures and focuses the explorative process of formulating hypotheses about what needs a concept answers to. Because we look beyond the concept's constitutive features for our standard

of appraisal, however, this process is more akin to hypothesis generation in the social and natural sciences than to deductive inference, and a need matrix will tend to be validated by its philosophical fruits rather than by how it was derived. In crystallizing for philosophical inspection a specific array of need-engendering conditions, however, a need matrix promises to render tractable the complexities of our predicaments: it gives us an uncluttered and perspicuous representation of a conceptual need and the combination of conditions engendering it.

Once a need matrix has been articulated, it can be probed with questions such as the following:

- *External Validation*: Do the conditions presented as generating a need for the concept in fact obtain? Is this true of one's own situation, now and around here?
- *Internal Validation*: Is the inference from their obtaining to there being a conceptual need for the concept valid? Does the concept really allow one to achieve something that could not be achieved without it or something very like it, given the circumstances and capacities envisaged?
- *Normative Endorsement*: Is the concern presented as motivating the use of the concept a concern one endorses, in the sense that one identifies with it and wants to see it satisfied?

In this way, explicitly setting out a need matrix can suggest questions we did not know to ask. It also has the virtue of holding up for scrutiny the key assumptions underpinning the claim that we need a certain concept. Probing whether the conditions presented as generating a need for the concept in fact obtain can provide an *external validation* of the model. Probing whether the inference from their obtaining to there being a conceptual need for the concept is valid can provide an *internal validation* of the model. And asking whether the concern presented as motivating the use of the concept is, upon critical reflection, a concern one wants to see satisfied can provide a *normative endorsement* of the model's evaluative basis.

A need matrix can thus be critically assessed much as other models are assessed. And, as with other models, its power lies in the handle it gives us on its target system, despite and indeed *because of* the simplification, abstraction, and idealization it involves:²⁷ it gives

²⁷ See Weisberg (2007, 2013), Strevens (2008, ch. 8), and Elliott-Graves and Weisberg (2014).

us something to work with, something we can explore, manipulate, put to the test, refine, enrich, extend, and elaborate.

Of course, achieving a remotely comprehensive understanding of the dauntingly intricate array of interacting concerns, capacities, and circumstances represented in a need matrix will require one to draw not merely on philosophy, but also on anthropology, ethnography, sociology, psychology, and history. These human sciences are better placed to capture the rich texture of the social reality in which our conceptual repertoire is put to work. Their findings can inform philosophers' model-building, both by schooling and sharpening philosophers' judgement as to what should figure in the model, and by offering empirical backing to the assumptions embodied in the model.

But precisely because of this division of labour, there is a point to distinguishing between the complex array of needfulness conditions we in fact face and the simple need matrix that professes to offer a selective philosophical representation of that array. The distinction between needfulness conditions and need matrices helps us to keep model and target system in their places and avoid conflations between them. A need matrix is not meant to be exhaustive. It is offered with a view to highlighting certain philosophically relevant aspects of our situation as part of a philosopher's case for or against a particular concept.

6.5 Needs-Based Concept Appraisal

Having introduced conceptual needs, needfulness conditions, and need matrices, we can formulate the following set of instructions for needs-based concept appraisal:

1. For any concept F , start by asking what kind of concept-user would have need of concept F : what combination of concerns, capacities, and circumstances would generate a conceptual need for F ?
2. Try to construct a need matrix—a representation of a minimal set of conditions jointly sufficient to generate a need for something very like F . The guiding constraint is that the matrix should be as specific to the concept as possible: not only should it generate a need for a class of concepts including F , but the class should be as small and tightly focused around F as possible, ensuring that the relevant need really is a need for F , or something very like it.

3. Once a need matrix calling for something very like *F* has been identified, consider whether the conditions actually obtain in your own case. Do you share that concern, those capacities and limitations, and the relevant circumstances? By coming to understand *whom* a concept would help to live, you achieve a better grasp of whether it actually helps *you* to live.
4. If you fail to share one or several of these conditions, the concept will to that extent be *pointless* for you (though its needfulness in other respects may be preserved by other conditions if the concept meets several needs at once), giving you a *pro tanto* reason to abandon it.
5. If, however, you share these conditions, the concept will to that extent be *pointful* for you, giving you a *pro tanto* reason to adopt it or adhere to it.
6. Having identified a need the concept answers to, you can then ask what demands this places on the concept, and whether there are respects in which the concept could be better tailored to that need.

As these instructions bring out, appraisal happens at two levels: first at the level of whether one should be using anything like this concept at all (is the concept pointless for me, or does it actually tie in with a need I share?), and then at the level of how well a given concept or conception is suited to meeting that need.

All six steps can then be reiterated to explore further conceptual needs the concept ties in with. A concept may have to strike a balance between several conceptual needs reflecting multiple and conflicting concerns.

In introducing the notion of a conceptual need, we focused on the analytically basic case, in which a single concern engenders a single conceptual need. But when we approach these issues from the rough-and-tumble of a concrete situation rather than in the abstract, we would expect—certainly on the kaleidoscopic picture—to find a turbulent confluence of different concerns, some pulling in diverging directions, each engendering its own conceptual needs.

The concept of a conceptual need therefore itself needs to be refined to meet this complexity. We need to distinguish between *pro tanto* conceptual needs, which one has *to the extent that* a particular need-engendering concern is to be satisfied, and *all-things-considered* conceptual needs. When a situation involves different concerns, we need to strike a balance between these concerns, and the concept we really need, all things

considered, will be the concept that best accommodates the various *pro tanto* conceptual needs growing out of these concerns. At the same time, we cannot arrive at an all-things-considered need without first considering the contributions made to it by individual concerns and the *pro tanto* needs engendered by them. One's all-things-considered conceptual needs are the resultant needs that emerge once a balance has been struck between one's different *pro tanto* conceptual needs.

Many of the concepts we use are, in practice, entangled in an overwhelming number of different concerns at any given time. What a need matrix aims to do is to render that complexity manageable for the purposes of philosophical reflection, at least to the point that we can think clearly, if only partially, about what kind of concept best serves our concerns. It will inevitably ignore various background concerns and perhaps leave out some important concerns as well, but this incompleteness is necessary, not just to thinking clearly about these matters, but to thinking about them at all.

Let us work through two brief illustrations of the needs-based approach to concept appraisal. Consider first the concept *poisonous*. What kind of concept-user would have need of such a concept? Here is a plausible need matrix for that concept, which isolates a set of needfulness conditions and highlights a resulting conceptual need:

Needfulness conditions: concept-users have an instrumental need for something very like the concept *poisonous* if and to the extent that the following conditions obtain:

- | | |
|-----------------|--|
| (Concerns) | They are concerned to avoid substances that have adverse effects on their bodies. |
| (Capacities) | They are incapable of tolerating a great many substances, but they are capable of exchanging information with similarly constituted concept-users. |
| (Circumstances) | They inhabit an environment in which there is a real risk of encountering substances with adverse effects on their bodies. |

Resulting conceptual need: concept-users satisfying all of these conditions have need of a concept that marks out things with adverse effects on their bodies as things to be avoided and enables them to alert each other to those things.

In view of this matrix highlighting a need the concept *poisonous* answers to, we can

consider whether the concept has a point for us, and what it needs to be like in order to serve us as well as it could. Clearly, we share all of the needfulness conditions singled out in this need matrix, which means that for us, the concept has a point, i.e. it meets a conceptual need of ours.

But the matrix also conveys a sense of what demands this need places on the concept and of what, in light of these demands, the concept *should* be like: the concept should have a descriptive dimension tracking certain biochemical facts; but the contours of its extension should also reflect a certain physiological perspective. It should track biochemical properties that have adverse effects on the human body, but not necessarily on other species. If some substance α turns out to have adverse effects on the human body, but does not fall under the extension of the concept as currently understood, then this is a respect in which the concept should be revised: the concept's extension should be broadened to include α . Furthermore, the concept should not be merely descriptive; it should carry a negative valence implying that one had better stay away from whatever falls under its extension. In this way, a need matrix allows us to appraise and improve a concept. Once the concept is in use, moreover, derivative forms of it tailored to other species can easily be created, as when an ethologist coins the concept *poisonous for the grasshopper mouse* and finds that scorpion venom, though poisonous in the original anthropocentric sense, does not fall under that derivative concept.

A second, particularly instructive example of needs-based appraisal can be constructed loosely on the model of Craig's 'synthesis' of the concept of knowledge, which—though Craig does not present it this way himself—can illuminatingly be elaborated into a needs-based reflection on the concept of knowledge we *should* have, given our conceptual needs.²⁸

We again start by asking who would have need of the concept of knowledge. Is there a human concern that, when conjoined with certain capacities and limitations under certain circumstances, would call for something like our concept of knowledge? Of course, the concept of knowledge is so deeply embedded in human life that it might be expected to tie in with several different conceptual needs at once; to take a fuller measure of its value, one would therefore have to reconstruct several different need matrices to

²⁸ See Craig (1986, 1990, 1993).

which it answers.²⁹ But Craig helpfully suggests that we start with one of the most basic concerns we can find that still displays some instrumental connection to the concept of knowledge.³⁰ He hypothesizes that an utterly basic concern that might entrain a need for something like the concept of knowledge is the concern, which we can hardly imagine human beings lacking, to gather information about the immediate environment, and especially about its risks and opportunities. If such a concern is felt in a community possessing a shared language, the mere fact that members of the community will be in different places at different times, and that no-one in the community will be capable of always acquiring all the information they need all by themselves, will be sufficient to generate a need for something like the concept of knowledge:

Needfulness conditions: concept-users have an instrumental need for something very like the concept *knowledge* if and to the extent that the following conditions obtain:

- (Concerns) *U*'s concern is to find out whether *p*;
- (Capacities) *U*'s own capacities are not sufficient to find out whether *p*, at least not without very considerable investigative effort;
- (Circumstances) *U*'s circumstances are such that there might be someone in *U*'s community who is in a position to tell *U* whether *p*.

Resulting conceptual need: concept-users satisfying all of these conditions have need of a concept that marks out *good informants* about whether *p*.

This concatenation of concerns, capacities, and circumstances characterizes what Craig dubs ‘the inquirer's situation’.³¹ In relation to the resulting conceptual need, a good

²⁹ Reconstructing several different need matrices to which the concept of knowledge answers is exactly what a lot of recent epistemological work can be read as having done. Other conceptual needs it has been thought to help us to meet include: the need to signal that inquiry is at an end (Kappel 2010; Kelp 2011; Rysiew 2012); to identify propositions treatable as reasons for acting (McGrath 2015); to provide assurance (Lawlor 2013); to distinguish blameless from blameworthy behaviour (Beebe 2012); and to honour subjects of knowledge attributions (Kusch 2009).

³⁰ See Craig (1990, 4).

³¹ See Craig (1990, 15). See also Fricker's (2007, 2010c) similarly needs-based commendation of the virtue of testimonial justice.

concept of knowledge will be a concept that is good *for* identifying good informants about whether p .

What kind of concept will be fit to meet the inquirer's need? The first answer that comes to mind is: a concept tracking the properties that are *constitutive* of being a good informant about whether p —which centrally include the propensity to say something true about whether p . But this will not do, because, for a concept-user confronting the need matrix of the inquirer as to whether p , this is no use at all in *recognizing* a good informant, since the ability to assess the truth of what someone says about whether p *presupposes* knowledge whether p , which is precisely the state that the inquirer hopes to attain by means of the concept of knowledge.

To be any help at all in meeting the conceptual needs of the inquirer, therefore, the concept of knowledge needs to track properties that *indicate* knowledge whether p and are *recognizable* to the inquirer as indicating knowledge whether p .³² These indicator properties might include: standing in the right causal relation to the state of affairs in question; being able to offer a justification for one's opinion; or having a good track record on this type of question. Each of these properties is *typically* a good indicator of knowledge, and the more such indicator properties a potential informant exhibits, the likelier it is that the inquirer is dealing with someone who actually knows whether p .

But the presence of these properties nonetheless falls short of guaranteeing that the inquirer is being offered knowledge whether p . A concept can be justifiably but nonetheless incorrectly applied if the presence of the indicator properties guiding its application merely makes it probable that the concept in fact applies.

This trade-off is nonetheless advantageous for the inquirer, since the choice is one between fallible conceptual guidance and none at all: the concern to find out whether p can be better satisfied by a concept that sacrifices infallibility to usability than by a concept that renders knowledge unrecognizable to those who do not already possess it. A concept of knowledge that could only justifiably be applied to people on the basis of their recognizably having a justified true belief about whether p might meet the conceptual needs of *examiners*, concerned to tell whether someone knows something already known to themselves, but it would do nothing to serve the concern to *find out*

³² See also Craig (2000, 656).

whether p .³³

The example illustrates that needs-based appraisal will not always favour the concepts that are the most reliable. The sort of concept of knowledge we need, as far as this need matrix allows us to see, is one that is less than fully reliable, but that makes up for its relative unreliability through its usability, by tracking properties that are readily accessible. Moreover, it should be a concept that also displays a considerable degree of *flexibility*: since the indicator properties through which a good informant becomes recognizable as such vary widely, and most of these properties are not strictly necessary conditions on being a good informant, we are better off using a concept that rides this variability, tracking the range of properties that typically but defeasibly indicate good informants whether p .

This simple need matrix modelling the inquirer's situation thus points to a reason to use *something* very like the concept of knowledge; and it even indicates reasons for us to use *certain forms* of the concept rather than others. As far as this need matrix allows us to see, the concept of knowledge we need should not single out any one indicator property by treating it as a necessary condition on the applicability of the concept. In view of the variability of the marks of a good informant, we do better to track a range of *normally* but not *unfailingly* reliable indicator properties without treating any one of them as indispensable. The need matrix thereby casts doubt on definitions of knowledge in terms of necessary and sufficient conditions. It presents the concept we need as being more flexible and less reliable for good reason.

6.6 Four Problems Solved

The needs-based approach to concept appraisal allows us to deal with the four problems we faced at the end of the previous chapter: the concern identification problem, the generalizability problem, the critical distance problem, and the underdetermination problem.

It solves the concern identification problem by exploiting the fact that, in *expressing* the conditions that render them needful, concepts guide us towards the concerns that might inform their appraisal.

³³ On the examiner's situation, see Williams (1973a, 146; 2005b, ch. 2).

It solves the underdetermination problem by highlighting how, if we bring certain concerns to the world, the world then makes certain demands on our concepts given what we demand of the world, thereby whittling down the space of possible concepts to those best suited to serving our concerns by meeting these demands. Only once we bring into the picture the practical demands that the world makes us on given our concerns does the notion of a good concept become fully determinate. This is why our concepts do not answer, in the first instance, to our concerns. In the first instance, they answer to our conceptual needs.

As for the generalizability and the critical distance problem, the needs-based approach overcomes them both by going more *external* than the goal-based approach we considered. On the goal-based approach, recall, ‘the concept’s epistemic goal is the kinds of inferences and explanations that the concept is intended to support’ (Brigandt 2010, 24), and ‘the epistemic goal pursued by a concept’s use is ... considered a component of a concept’ (Brigandt 2010, 22). Needs-based appraisal, by contrast, looks to how concepts directly or indirectly tie in with people’s concerns by meeting their conceptual needs, independently of whether these concerns are constitutive of the concepts, or of whether people use the concepts with a view to satisfying these concerns. Accordingly, while constitutive goals are internal to concept-users’ understanding of what they are up to in using concepts, their conceptual needs can come as a discovery to them.

This needs-based standard is more external also in that many of our concerns are not subject to our will in the way that our goals or aims are. What concepts we need can therefore be independent of our will. It is only in special cases, namely those in which all the relevant (i.e. need-engendering) concerns are fully under the control of the will, that we can change our conceptual needs at will, by *redirecting* our concerns. But this only works when we can redirect our concerns by *deciding* to redirect our concerns. And many of our most fundamental concerns, including those growing out of what we have called our inner needs, are concerns for things that we cannot simply decide not to care about. Concept-users might just have certain conceptual needs in virtue of their concerns and limited capacities in the kind of world they live in, whether they want it or not. In sum, conceptual needs form a more external standard than constitutive goals in that needs can be had unwittingly and unwillingly.

Such greater externality brings both advantages and disadvantages. The main

disadvantage, which became evident already in the examples we considered, is that constructing a need matrix requires a good deal of interpretation: one cannot mechanically read off the concept what conditions should figure in the need matrix; rather, one needs to searchingly work one's way back to some human concern, tentatively combine it with certain capacities and circumstances so as to obtain a set of need-generating conditions, and explore how much light this sheds on the concept. As in the hypothetico-deductive model of science, these initially tentative conjectures can come to be affirmed by their consequences and the consonance of the overall picture they lead to. But they thereby inevitably also risk seeming *ad hoc* at first, especially if compared to claims about the concept that purport simply to follow deductively from the concept's constitutive features.

Yet claims about a concept's constitutive features and what follows from them can seem contentious and *ad hoc* as well—especially if they concern the concept's constitutive goals, since concept-users may not agree what the goal is, or only possess an inchoate understanding of it that makes any particular articulation of what the goal is look controversial.

And, above all, to say that claims about what needs a concept ties in with require interpretation is not to say that anything goes. We saw that need matrices can be externally validated, internally validated, and normatively endorsed. Furthermore, the way in which Craig's account is being scrutinized, refined, complemented, and improved on by other philosophers exemplifies the validating and corrective role of the court of philosophical opinion.³⁴ In the end, whatever interpretation is involved in the needs-based approach is constrained by what genuinely makes sense to us and what remains obstinately unintelligible, what holds up and what falls flat. In the light of a concept's features and the way it is in fact employed in human life, some claims about the needs it answers to will make more sense than others, and some will simply not make any sense. There are claims about conceptual needs that no truthful look at human affairs will bear out. It stands with claims about conceptual needs much as it stands with interpretative claims in disciplines like evolutionary biology, archaeology, sociology, or history. There is room for interpretation, but there are also claims that no truthful reading of the data

³⁴ See Williams (2002; 2010), Fricker (1998, 2007, 2012, 2016a), Kusch (2009, 2011, 2013), Pritchard (2012), Reynolds (2017), and Hannon (2013, 2015, 2019).

will bear out—as the French Prime Minister Clemenceau replied to a representative of the Weimar Republic wondering how future historians would make sense of the outbreak of WWI: ‘This I don’t know. But I know for certain that they will not say Belgium invaded Germany’.³⁵

Interpreting one’s way to a more external standard also has significant advantages, however, in that it solves the problems of generalizability and critical distance.

Consider first the critical distance problem. By stepping back from a concept and looking at its relation to our conceptual needs, we obtain critical distance not only towards the concept, but also towards its constitutive goal, and potentially even towards the aim that concept-users consciously pursue in using the concept. Concepts then invite critique to the extent that they fail to adequately meet the conceptual needs that concept-users have in virtue of their concerns, capacities, and circumstances. This makes room for the possibility of finding, on the basis of our conceptual needs, that even a concept that is perfectly suited to meeting its constitutive goal, or to realizing our conscious aims, nonetheless fails to meet our conceptual needs, and, at worst, systematically obstructs or frustrates them.

But needs-based appraisal also goes beyond goal-based appraisal in allowing for the possibility that concepts might serve our concerns even though we do not regard them as instrumental to our concerns at all. When philosophers draw their standard of concept appraisal from the instrumentality of a concept in serving its constitutive goal, they have no use for the distinction between the perspective of the engaged concept-user and the perspective of the disengaged, reflective appraiser of the concept. Nor does any approach that takes the conscious aims or goals with a view to which a concept is employed as its standard. For when a concept is appraised according to how conducive it is to meeting the goal with which it is deployed, user and appraiser alike explicitly *instrumentalize* the concept by regarding it as a means to achieve a certain end, and so the instrumentalizing mentality of the appraisal is in keeping with the instrumentalizing mentality with which the concept is employed by its users.

By contrast, needs-based appraisal allows for the possibility that our concepts might possess instrumental value we know nothing of, as they may fill needs we did not even

³⁵ The exchange is related in Arendt (1968, 239). For a discussion of constraints on historical sense-making, see Williams (2002, 241–50) and Cueni and Queloz (2022).

know we had. Seemingly idle concepts may in fact do important work; concepts of intrinsically valuable traits or properties may be essential props of practices and institutions without being used in an instrumental spirit at all. The *classical gene* concept is a special case, in that there is some plausibility to the claim that it was consciously designed and deployed to serve scientists' aim of explaining and predicting inheritance patterns. This instrumentalizing mentality obligingly hands us a standard for appraising the instrumentality of the concept. But many branches of human thought—esthetic, moral, political, or legal—are thick with concepts that are not evidently goal-directed in this way, and hence provide nothing like the focused evaluative guidance of the Mendelian geneticist's naked instrumentalism about his purpose-built concepts. We can still ask how a concept lacking a constitutive goal relates to our conceptual needs and try to gauge the concept's instrumental value to us.³⁶ But the concept itself will not always hand us the standard for this appraisal. We may have to look beyond the concept and reconstruct what kinds of exigencies, if any, it might be responding to, because its instrumental dimension is initially opaque. Part of what lends needs-based concept appraisal its philosophical interest and informativeness, then, is that our conceptual needs are not immediately transparent to us in the way that the goals with which we use certain concepts are.

Making room for critical distance from our concepts and the goals with which we employ them is important not just to account for the possibility of radical critiques of inherited ways of thinking, but also, more generally, for separating what is dead from what is alive in our conceptual inheritance. Some concepts can be shown by reflection on our conceptual needs to have outlived their usefulness—they may have been perfectly good concepts for those who first created them, but they are dead wood to us. Understanding a concept's needfulness conditions equips us to discriminate between the conceptual holdover that has outlived its usefulness and the still indispensable concept, the otiose and the continually valuable.

Consider, for example, the concept *honour*. Who would have need of such a concept? Drawing on work from social and political science, we might model its needfulness conditions using the following need matrix (which, by definition, is meant to be no more

³⁶ For an approach to esthetic concepts that is highly congenial to the needs-based approach, see Robson and Sinclair (2022).

than a simplified and partial representation of the wider array of a concept's needfulness conditions):³⁷

Needfulness conditions: concept-users have an instrumental need for something very like the concept *honour* if and to the extent that the following conditions obtain:

- (Concerns) Individuals are concerned to hold on to their property and deter theft.
- (Capacities) Individuals' capacity to prevent theft is limited—especially when it comes to highly divisible and portable property.
- (Circumstances) Centralized institutions enforcing property rights are weak or entirely lacking.

Resulting conceptual need: concept-users satisfying all of these conditions have need of a concept that compensates for the lack of centralized institutions enforcing property rights by introducing a more diffuse and decentralized deterrent.

In relation to this need matrix, the concept of honour might be thought to serve property-owners' concerns, whether they realize it or not. For it can act as the linchpin of an honour culture in which people are quick to take offence and can be expected to retaliate even when the costs of retaliating exceed the value of the stolen good. And by projecting a willingness to treat even a comparatively small theft as a serious offence calling for retaliation *in the name of honour*, people send discouraging signal to potential thieves, compensating the lack of a centralized deterrent with a more diffuse and decentralized deterrent.³⁸

This need matrix could be used to account for the prevalence of the concept of honour in places where its needfulness conditions were fulfilled—it could help explain the strong honour cultures in places like the American South in the eighteenth and early nineteenth century, where valuable property tended to take the form of cattle rather than land, and centralized institutions tasked with enforcing property rights were comparatively weak

³⁷ See especially Testini (2021a, b), but also Nisbett and Cohen (1996) and Shackelford (2005). There is of course more to be said about the needs met by the concept of honour: its role as an engine of moral reform, for instance, is described in Appiah (2010).

³⁸ See Testini (2021a, b).

or unreliable.

But equally, the need matrix can be used to assess the value of the concept for us today. Insofar as we fail to share one or several of the three central planks of the need matrix, we will lack reason to think in terms of *honour*. For us, the concept will to that extent be pointless. This illustrates how the needs-based approach allows us both to understand why a concept earned its keep under certain conditions while at the same time being unhelpful *to us*, for whom these conditions no longer hold. Reconstructing the concept's needfulness conditions has critical force for us—and does so precisely because the conditions it appeals to are *not* constitutive of the concept, for this is what allows the concept to *outlive* the conditions that formerly bestowed value upon it.

Secondly, going more external also enables the needs-based approach to overcome the generalizability problem: even where a concept lacks a constitutive goal, or where we do not consciously deploy it with a view to satisfying a particular aim, we can still look to how the concept relates to our conceptual needs.

The greater generality of the approach does not consist merely in its wider applicability to our current concepts. While looking at the fits between concepts and need matrices allows us to evaluate how well the concepts we now use meet the conceptual needs we now have, the same relation of fit between concepts and need matrices can be deployed *retrospectively*, to discern the respects in which, even when concepts lacked a constitutive goal, conceptual change was nevertheless rational and not just the contingent product of brute causal forces: it was a rational adjustment to corresponding changes in the needfulness conditions to which the concepts answered.

Moreover, this relation of fit between concepts and need matrices can also be deployed *prospectively*, to get a sense of what concepts we will need as we meet the future and the novel concerns, capacities, and circumstances it brings. If our concerns, capacities, and circumstances change, our conceptual repertoire will have to change with them if it is to meet our conceptual needs. Developments such as climate change and the increasing reliance on AI bring unprecedented conceptual needs in their wake—perhaps we need to fashion many more concepts of 'green virtues' to extend our repertoire of inherited virtue concepts; or perhaps we need to rethink our conceptions of liability for

clinical decision-making relying on opaque AI-powered systems.³⁹ Constructing need matrices can help us envision these needs and what the concepts capable of meeting them should look like.

And, lastly, looking to conceptual needs also gives us the ability to discriminate between conceptual engineering efforts that fill a need and conceptual engineering efforts that are needless or ill-conceived. As the American philosopher Justus Buchler succinctly concluded: ‘the responsible introduction or extension of terms, whether in philosophy or science, reflects a conceptual need’ (1955, 108).

³⁹ On ‘green virtues’, see Jamieson (2007, 2014); for an overview of the issues surrounding the use of AI in clinical decision-making, see Smith (2021).

CHAPTER SEVEN

Reasons for Concept Use

In search of a framework within which to answer the authority question, we took from the Dworkin–Williams debate the idea that our concepts should tie in with the concerns we identify with, because conceptual authority ultimately wells up from our concerns. But we then saw that, even if our concepts must tap into our concerns to be imbued with authority, these concerns do not yet yield an operational measure of a concept's merits; nor can concerns by themselves give us reasons to use certain concepts rather than others. That is done by the conceptual needs our concerns generate through their interaction with our capacities and circumstances. It is in our conceptual needs—many of which are highly local—that we find reasons for concept use.

My aim in this chapter is to clarify and circumscribe the role of reasons for concept use before considering how they allow us to answer the authority question. Along the way, I fend off several looming misunderstandings to bring out how we can allocate a role to these reasons in metaconceptual reflection on how concepts relate to our concerns without thereby crowding out other kinds of reasons that transcend preoccupation with our concerns or with the instrumentality of concepts. In fact, it will emerge that reasons for concept use can boost our confidence in ways of thinking that are not centred on human concerns at all.

7.1 Reasons in vs. Reasons for Concept Use

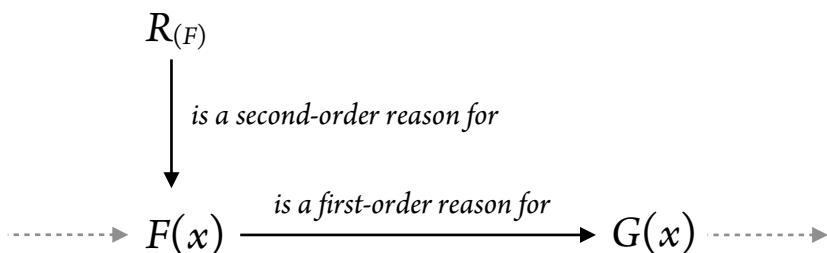
When one steps back to assess the authority of a concept, one disengages oneself from

the concept and its correlative reasons. The pattern of reasoning traced out by the concept is put in the dock and critically evaluated: what reasons do we have to heed those reasons in the first place? From this more external vantage point, we can then look to the concerns we identify with and consider what conceptual needs they engender in us, given our capacities and circumstances. These needs yield reasons to use certain concepts and be responsive to certain reasons rather than others.

When a justification such as ‘Because it is *natural*’ prompts the authority question and leads us to demand reasons to recognize this as a reason, we can accordingly work our way up the chain of reasons to determine what reasons for concept use the authority of this concept and its concomitant reasons is grounded in.

Yet this is to work one’s way up a chain of reasons that is orthogonal to the chains of reasons we move along as engaged users of a concept. Ordinarily, we move along the chain of reasons constituted by object-level judgements, asking what a judgement follows from, or what follows from it. In the example of the judgement articulated in terms of *naturalness*, a normal request for reasons would take the form: ‘Why is it natural?’ That question takes the concept *natural* for granted and demands a justification for this particular application of it.

The authority question, by contrast, involves an ascent to the metaconceptual level of second-order reasons, where it asks for reasons to regard its being natural as a reason in this connection. This is to demand reasons for reasons, i.e. second-order reasons to use a given concept *F* such that *x*’s being *F* counts as a first-order reason to think that *x* is *G*:



A reason for a reason: $R_{(F)}$ is a second-order reason to use a concept *F* such that *x*’s being *F* counts as a first-order reason to think that *x* is *G*.

We can sharpen the contrast between, on the one hand, the first-order reasons *in terms of which we think* when using concepts in an engaged way, and, on the other hand, the

second-order reasons *for thinking in these terms* that come into view when using the concepts in a disengaged way, by distinguishing between reasons *in* and reasons *for* concept use:

Reasons in concept use are the ordinary reasons operative within a practice of reason-giving, i.e. the reasons guiding and flowing from the application of concept F that one has insofar as one is an engaged user of concept F . From a disengaged perspective on concept F , a reason in concept use might register as follows: ‘For an engaged user of the concept F , $F(x)$ is a reason to think that $G(x)$ ’. From an engaged perspective on concept F , by contrast, the same reason would register simply as: ‘ $F(x)$ is a reason to think that $G(x)$ ’.

Reasons for concept use are the reasons for a practice of reason-giving, i.e. the second-order reasons one has to be an engaged user of a concept F such that one is disposed to treat $E(x)$ as a reason to think that $F(x)$, or $F(x)$ as a reason to think that $G(x)$. Let $R_{(F)}$ be such a second-order reason to use concept F . From a disengaged perspective on concept F , $R_{(F)}$ must not itself make use of concept F . From an engaged perspective on concept F , by contrast, no such restriction holds on what can count as $R_{(F)}$.

The two kinds of reasons are connected as follows: reasons *for* concept use are reasons for recognizing certain reasons *in* concept use *as* reasons—in this sense, they are reasons for reasons.

By way of illustration, consider again the example of the concept *blasphemous*: the reasons *in* concept use are the reasons that guide and flow from the application of the concept—reasons that might figure in deliberation as follows: ‘This book is *blasphemous*, because it is disrespectful towards God’; or ‘Because the book is *blasphemous*, it should be banned’. Reasons *for* concept use, by contrast, are the rationales that underlie or vindicate the adoption and continued use of the concept—they are the reasons one has to reason in terms of the concept *blasphemous* and heed the reasons it adverts to.

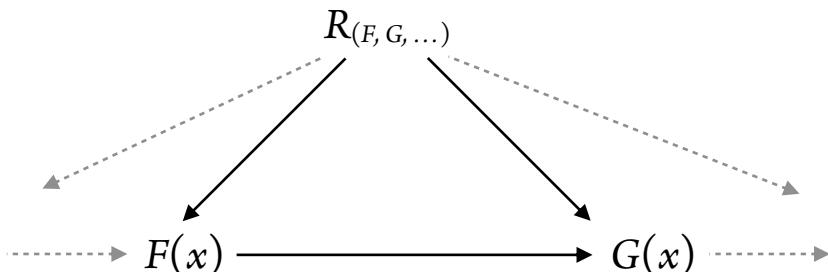
To be compelling in the eyes of someone who is not fully confident in the concept, reasons for concept use must not themselves draw on the concept whose use they favour: in this example, one *could* insist that we have need of a concept of blasphemy because

there is so much blasphemy around, and because what is disrespectful towards God should be banned; but this would be not so much *false* as *insufficiently independent* of the concept and its concomitant reasons. Precisely what someone challenging the authority of the concept wants to know is whether we have reason to be sensitive to the presence of blasphemy and heed the reasons associated with it. To be forceful independently of one's confidence in the concept, therefore, reasons for concept use will have to be reasons given from a disengaged perspective on the concept. Of course, those reasons for concept use will still make engaged use of other concepts—inevitably, *all* reasons are reasons revealed *in* the use of some concept or other (this is what led Wittgenstein to worry that giving reasons for thinking as we do would be impossible if required an answer 'outside the game of reasoning' (1979, §4)). But no problematic circularity is involved in this as long as the reasons revealed in the use of some concepts offer reasons for the use of other concepts.

In its commitment to the idea that it is worth asking, of any practice of reason-giving, whether there is some reason for it, the needs-based approach remains moderately rationalistic. It contrasts in this respect with a view on which concept choice is always rationally contingent and a mere matter of convention. The needs-based approach takes seriously the possibility that there might be *some reason for each* practice of reason-giving.

But it equally contrasts with views on which there is one reason, or one unified set of reasons, for all reason-giving practices. Both foundationalist attempts to derive the right set of concepts from timeless rational foundations and tidy-minded attempts to derive them from a unifying theory embody the rationalistic hope that there might be *one type of reason for all* reason-giving practices, for instance in the form of a universal metric of

utility, or of a categorical imperative:



One reason for all reasons: $R_{(F, G, \dots)}$ is a second-order reason to use a concept F such that x 's being F counts as a first-order reason to think that x is G , as well as a second-order reason to use a concept G such that ..., etc.

It is this ambition to find a unified basis of authority that the needs-based approach abandons, looking for a form of conceptual authority that is grounded in a plurality of needs and concerns instead of being derived from a single overarching principle.

The needs-based approach still seeks to identify bases of authority in the various concerns of concept-users and their resulting conceptual needs that allow reasoned discrimination between more and less authoritative concepts; it just gives us no reason to expect there to be a single currency of reasons that everywhere forms the basis of authority. As Anscombe remarks, we can accept that all chains of reasons must stop somewhere, but we should resist the ‘illicit transition’ from ‘all chains must stop somewhere’ to ‘there is somewhere where all chains must stop’ (1957, §21).

On this view, there are as many different types of bases of authority as there are types of human concerns. Only together with the conviction that there must be a single currency of reasons for all reasons does the Enlightenment expectation that we can always appropriately demand reasons for reasons culminate in a unified system.¹ Without this conviction, we can seek reasons for each but not for all reasons, and we will then end up not with a unified system, but with a pluralistically vindicated conceptual apparatus:

¹ The ‘single currency of reasons’ is Williams’s phrase; see Williams (2011, 250n13). See also Williams (2011, 124–6) for a related discussion of how the demand for what Williams calls ‘justificatory reasons’ naturally leads to the construction of an ethical theory.



Reasons for each reason: $R_{(F)}$ is a second-order reason to use a concept F such that x 's being F counts as a first-order reason to think that x is G . $R_{(G)}$ does the same for the concept G , etc.

Such an approach may not live up to the demands of tidy-mindedness, but it nevertheless promises to meet two desiderata on a non-foundationalist understanding of conceptual authority. It promises to enable us (i) to identify reasons to be confident in at least some of our concepts, thus avoiding indiscriminate ironism; and (ii) to make sense of the possibility of radical critique of our concepts in terms of the possibility that the concepts currently in use fail to meet our conceptual needs, thus avoiding undiscriminating holism.

7.2 Concern-Independent Reasons in Concept Use

The idea that conceptual needs constitute reasons for concept use capable of buttressing reasons for action and belief must not be allowed to displace a more entrenched fact, however: that, for the engaged concept-user, there are countless places in the conceptual architecture where the chain of reasons comes to an end in buck-stopping, spade-turning reasons that have nothing whatsoever to do with our needs or concerns.² It would, again, be a conflation of the engaged and the disengaged perspective to conclude, from the fact that reasons for concept use are to be found in conceptual needs deriving from human concerns, that all chains of reasons must end in conceptual needs or in human concerns.

For a needs-based account of concept appraisal to be psychologically realistic and steer clear of reductive oversimplification, therefore, it is crucial to recognize how *independent* of our concerns the reasons in concept use buttressed by reasons for concept

² A point recently emphasized, from within frameworks congenial to the needs-based approach, by Müller (2020, 189–218) and Sinclair (2021, 191–217)

use can be. Except in certain cases—a concern to look different can be cited as a reason to get a different haircut—the reasons that move us to believe or do things frequently make no reference to our concerns at all.

It is a remarkable and important fact about concepts that they have the power to direct our attention away not just from an ego-centric preoccupation with the satisfaction of our selfish concerns, but even from an anthropocentric preoccupation with the satisfaction of human concerns. They enable us to discern completely *concern-independent reasons* in certain objective facts and properties. Some concepts enable one to discern forms of value in states of the world that do not even have any human beings in it, let alone human concerns. Environmental values such as the value of biodiversity, for example, need not, at the level of their conceptual content, be centred on human beings.³ Likewise, many of our epistemological concepts—of *truth*, *belief*, and *evidence*, for example—enjoin us to consider it an impropriety to let practical considerations of concern-satisfaction count as evidence for the truth of a belief. When it comes to assessing the truth of a belief, the fact that the belief might serve my concern for happiness is rightly thought to be neither here nor there. These concepts make a point of insisting that the first-order reasons involved in living by them are splendidly indifferent to human concerns.

Nothing in the present account gives us reason to deny or dilute the disregard for practical considerations and human concerns encoded in many of our concepts. Quite the opposite: from a disengaged perspective on these concepts, we may well find that there are compelling second-order reasons for these concepts to be so picky in what they will permit us to count as first-order reasons.

A notable virtue of the distinction between reasons in and reasons for concept use is that it allows us to reconcile and do justice to seemingly incompatible aspects of reasons. On the one hand, *internalist* or *subjectivist* accounts of reasons have emphasized the respects in which one's reasons seem to be a function of one's subjective concerns. On the other hand, *externalist* or *objectivist* accounts of reasons have emphasized the respects in which there seem to be objective reasons whose force is precisely not a function of one's subjective concerns. Using the distinction between reasons in and reasons for concept use, we can accommodate and reconcile the intuitions of both sides.

³ See Williams (1995f) and Krebs (1999).

The objectivist is right to insist that many of the reasons that properly inform our deliberation—reasons *in concept use*—must be understood as being completely independent of our concerns (or, as the Williamsian terminology that started off these debates has it, our ‘motivational sets’).⁴ These reasons are not a matter of what we happen to care about, but of what objectively matters, whether we care about it or not. They are reasons flowing from objective facts and duties, or from the objective goodness or badness of things. From the perspective of the engaged concept-user who is applying the concept and considering the reasons it adverts to, any further considerations as to how the use of the concept ties in with people’s concerns can seem like ‘one thought too many’.⁵ It jars with the perspective opened up by the concept.

And yet, from the disengaged perspective of one who reflects on what reasons they have to think of the world along these reason-giving lines in the first place, the blank refusal to contemplate anything other than the reasons the concept immediately adverts to can seem, as Simon Blackburn puts it, like ‘one thought too few’:

a lawyer who becomes convinced that his branch of law exists in order to protect extant distributions of property, and who begins to think that those distributions are themselves disastrous, may reasonably lose his enthusiasm for the rules of law that he has been trained to enforce. But ... it is natural and good that disenchantment with the consequences should feed back into disenchantment with the institutions and the rules that exist to promote them. A deontologist who is insulated from this feedback is, like the man who starves for want of a dinner jacket, a lunatic rather than a saint. He has the converse vice of the man who has ‘one thought too many’, namely, that of having one thought too few. (1998, 45)

The person who is myopically bound up in the application of concepts serving concerns the person would disapprove of upon reflection *should* be more responsive to metaconceptual considerations about the concept’s role in the larger scheme of things. Blind confidence in a concept can be as negligent as the blind following of rules and orders.

This is where we find a place for the subjectivist thought that the reasons people have to conceptualize things along certain reason-giving lines rather than others depend

⁴ See Williams (1981b).

⁵ See Williams (1981e, 18).

notably on how these conceptualizations relate to their concerns. Even when the reasons that guide our thought and conduct make no mention of our concerns, these concerns nevertheless play a role in engendering the conceptual needs that give us second-order reasons to let ourselves be guided by certain concern-independent reasons. Hence, while the application of a concept may properly be responsive only to reasons that are *not* a function of concept-users' concerns, the metaconceptual appraisal of a concept may equally properly be responsive to reasons for concept use that *are* a function of concept-users' concerns. Even where we reason in concern-independent ways, concerns properly figure as an explicit and constantly relevant parameter at the higher-order level of the reasons for us to reason in concern-independent ways.

Acknowledging that engaged concept use answers to different kinds of reasons than disengaged concept use therefore allows us to find an appropriate place for subjectivist as well as for objectivist intuitions. As Blackburn's example of the lawyer intimated, moreover, we thereby also accommodate consequentialist intuitions, which are responsive to how the effects of using a concept affect our concerns, as well as deontological intuitions, which are responsive to the fact that many of our concepts precisely enjoin us to disregard such considerations. We can grant both that disengaged metaconceptual reflection appropriately looks to how the consequences of concept use tie in with human concerns, and that, except in certain cases, the reasons that figure in one's engaged, object-level deliberations often make no reference to concerns at all, and demand to be understood as being completely independent of them: they are objective and external considerations, a matter of duty and what is right, irrespective of the concerns people happen to have. In the Wittgensteinian analogy, it can be an essential part of a language game to deny that its subject matter depends in any way on our language game. The reasons in concept use adverted to by many of our concepts present themselves and are correctly understood as *concern-independent* reasons: the fact that something is unjust, or robs someone of their liberty, figures in the deliberation of the engaged concept-user as a self-standing reason, a reason that is simply there, regardless of whether the people involved have certain concerns or not.

7.3 Instrumentality Without Instrumental Mentality

Having distinguished reasons for and reasons in concept use and underscored the

importance of keeping them in their places when switching between engaged and disengaged use, we also need to forestall a closely related misunderstanding of the needs-based approach that it is crucial to avoid: in approaching concepts from our concerns and the instrumental needs they engender, it is important not to conflate the instrumentality of thought with *instrumental thought*. If we conflate the two, corrosive instrumentalism threatens—an instrumentalizing mentality undermining all rational authority that is not the authority of its own concerns. It is an old threat, to which the ancient Greeks were already acutely alive.⁶ To dodge it, we must find a place for the instrumentality of concepts in metaconceptual reflection while taking care not to collapse all rationality into instrumental rationality, or all reasons into instrumental reasons.⁷

While it is true that concepts are needed, when they are, as a matter of technological necessity, it is plainly not true that all our thought is technological in character, i.e. embodies an instrumentalizing mentality at the level of its content. Using a certain concept might be instrumental to meeting one's concerns, but it does not follow that the concept used casts its object in an instrumental light. On the contrary, it may be that using that concept is instrumental only insofar as the concept casts its object in a non-instrumental light. Thinking in terms of certain concepts can be instrumental to the satisfaction of our concerns even when—indeed, just because—it is not instrumental-minded.

There are many things that are not best attained by consciously striving for them. Various feats of technical prowess are not best achieved by consciously thinking of achieving them, and it is a philosophical commonplace that happiness is not best pursued under that description. Game theory offers a slew of more rigorously formalized illustrations:⁸ the chicken game, the free rider problem, the ultimatum game, the stag hunt, the centipede game—they all involve the idea that simply thinking in terms of the basic concerns in terms of which the game is set up will lead one to lose the game. And they all encourage the conclusion that one salient type of solution to this difficulty can

⁶ On the threat of instrumentalism in ancient Greek thought, see Ober (2022, 2, 362–72).

⁷ For a disambiguation of the different implications that such a collapse might be taken to have, see Kolodny and Brunero (2020, §4.1).

⁸ See Skyrms (1996, 2004), Binmore (2005, 2007), and Kitcher (2011) for discussions that remain particularly germane to philosophical concerns. See also Bowles and Gintis (2011).

be found at the level of conceptual innovation: the structure of the game gives one reason to adopt a concept—the concept of pride, perhaps, or the concept of loyalty, or the concept of honour, or the concept of a promise—which, by rendering one responsive to new reasons that are precisely *not* a function of the basic concerns structuring the game, leads one to satisfy those basic concerns *better* than one otherwise would. Each of these game-theoretical structures indicates a way in which a concept can help satisfy a concern by distracting from that concern.

Accordingly, there can be instrumentality without instrumentalizing mentality. A concept can be a device serving a certain concern without encoding a preoccupation with that concern. The intrinsic value of truth, for example, is an intrinsic value precisely because it is not a matter of serving antecedent human concerns—the value of finding out the truth and the disvalue of lying are not simply a function of how many concerns get satisfied or frustrated as a result, and one would misunderstand the value of truth if one understood it entirely in such instrumental terms. Nevertheless, there is a good reason why creatures like us have the concept of a value that is concern-independent in this way: it notably serves the constant human concern to acquire information about the risks and opportunities of one's environment by enabling an epistemic division of labour whereby members of a community cooperatively *pool* information.⁹ Crucially, however, the concept of the value of truth can only properly serve this concern if it is not conditional on subservience to human concerns. For it is only *insofar* as the truth is regarded as *intrinsically* valuable that valuing the truth can be *instrumentally* valuable. As long as individuals value the truth only *insofar* as it serves their individual concerns, the concept will fail to make a helpful difference: people will try to free ride on the pool of information without themselves bothering with the truth unless they immediately stand to gain from doing so anyway. And since free riders not only do little to enrich the pool with hard-to-get information, but vitiate it with misinformation whenever they can profit from misleading people, the epistemic division of labour will be terminally unstable.

For the concept of the value of truth to serve human concerns by sustaining the epistemic division of labour, therefore, it needs to be the concept of something valued for its own sake, which throws its own weight into the balance of reasons, so that people seek out the truth and tell it to others *just because* it is the truth. Once considered intrinsically

⁹ See Williams (2002). I reconstruct this explanation in detail in Queloz (2018; 2021b, ch. 7).

valuable, the truth becomes its own reward, and can stake a claim against narrow self-interest. This may not suffice to override self-interest every time, but it enables a sustained practice of information pooling to get off the ground, and the occasional lie will not bring it down.

The value of truth, on this account, needs to outrun its instrumentality in order to be instrumental to our concerns. But our ability to grasp its instrumentality turns, once again, on disambiguating the ‘ing/ed ambiguity’: our *valuing* the truth intrinsically, regardless of human concerns, is instrumentally valuable in meeting human concerns, but we can only grasp how it can be so if we grasp why *what is valued* is not valued merely instrumentally, as a means of meeting human concerns. P. F. Strawson may have been getting at a similar point when he suggested that we can only really understand the efficacy of some of our reactive attitudes if we recognize that they express moral and not just instrumental attitudes.¹⁰ The point is not just that they combine instrumental and non-instrumental aspects. The point is that the instrumental aspects *depend* on the non-instrumental aspects. We are bloody-minded rather than benefit-minded in blaming others, for instance, because blame only serves our concerns to regulate anti-social behaviour and align our moral sensibilities on the condition that it be understood as justified by something other than its subservience to such concerns.¹¹

Similarly, the authority of many concepts and ideals whose superb indifference to mundane concerns is prone to arouse suspicions in the wake of Enlightenment naturalism, rationalism, and humanism might be vindicated by the realization that there are good reasons for this indifference: we sometimes think in concern-independent terms the better to serve our concerns. This is of course not to deny that many concepts do not fall into this category: we sometimes also think in concern-dependent terms the better to serve our concerns, or would be better off if we did so. Moreover, a concept might not serve *our* concerns at all: it might serve someone’s else concerns at our expense. And finally, a concept might of course also not serve anybody’s concerns, and simply be idle.

Yet, in the present context, the most philosophically interesting and relevant case is

¹⁰ See Strawson (2008a, 27).

¹¹ I elaborate on this argument in Queloz (2021c). On the connection to Strawson, see also Emilsson (manuscript).

that in which we think in concern-independent terms the better to serve our concerns. The application of a concept will then yield and be guided by reasons that are completely emancipated from our concerns. But what gives this concept its importance is its relation to our concerns. It thus stands with concepts much as Wittgenstein says it does with rules: ‘A rule *qua* rule is detached, it stands as it were alone in its glory; although what gives it importance is the facts of daily experience’ (1978, VII, §3). He goes on to compare the task of describing this combination of features to the task of describing the office of a king: one must avoid the error of reducing the dignity of that office to its usefulness, yet leave neither its dignity nor its usefulness out of account.¹² What I am suggesting is that this can be achieved by bringing out the usefulness of the king’s office possessing a dignity that is precisely not grounded in its usefulness, but independent of it. Likewise, to make sense of the fairly pervasive phenomenon whereby a concept has to emancipate itself from human concerns in order better to serve them, we must avoid reducing the autonomous content of a concept to its instrumentality, while at the same time leaving neither its autonomy nor its instrumentality out of account.

The picture painted by the needs-based approach, then, is one on which reasons for concept use are *bona fide* reasons that our confidence in concepts should be responsive to; but they are reasons that do not directly compete with, and therefore should not be at risk of crowding out, reasons in concept use.

To adopt a needs-based approach to the authority of concepts is thus not to fall for the reductive misconception that all concepts which serve human concerns do so by introducing a preoccupation with concern-satisfaction into the content of people’s deliberations. On the contrary, one cannot fully appreciate the power of the conceptual architecture we inhabit, or indeed its instrumentality, unless one recognizes that many concepts serve us best by transcending and screening out all considerations as to how their use relates to human concerns.

7.4 From Concerns to Reasons in Concept Use

And yet, the mere fact that many concepts serve us best by screening out all considerations as to how their use ties in with our concerns does not mean that these

¹² The Bagehottian resonances of this example are explored in Bloor (2004, 125–6).

concerns do not play an essential role in rendering these concepts needful and pointful—just as the fact that only the actors actually appear in the play does not mean that the people working backstage do not play an essential role in the theatre.¹³

Even when the reasons that guide our thoughts and actions make no mention of our concerns, these concerns nevertheless give us second-order reasons to deliberate in terms of certain concepts rather than others, thereby determining what reasons to think or do certain things we have reason to be guided by. This is where we find an indirect conduit of authority from concerns via needs to concept use and its concomitant reasons. This indirect conduit looks as follow:

- concerns, when pursued with certain capacities under certain circumstances, yield *conceptual needs*: needs for certain concepts rather than others;
- conceptual needs yield *reasons for concept use*: reasons to use concepts that are tailored to those conceptual needs;
- concept use yields *reasons for concept application*: when something fulfils the criteria guiding a concept's application, this gives one reason to apply the concept to it;
- concept application in turn yields *reasons for belief or action*: when one applies a concept to something, this gives one reasons to draw certain consequences of the concept's applicability in one's thought and action.¹⁴

Our concerns and the reasons for action and belief that immediately guide our thought and conduct are therefore linked, albeit highly indirectly. Take, for example, the concept *fire*. Our concern to stay out of harm's way, when combined with a combustible environment and the human body's incapacity to tolerate even brief and local exposure to temperatures above a certain threshold, creates a conceptual need for something like the concept *fire*, which in turns gives us a reason to use the concept. To be an engaged user of a concept of fire worth having is to be sensitive to the fact that the presence of certain conditions (i.e. heat, light, smoke, and flames) yields reasons, however defeasible, for the application of the concept *fire*. From the applicability of the concept, it in turn had

¹³ Wittgenstein (1978, V, §15) draws a similar comparison in discussing the role of empirical facts in mathematics.

¹⁴ Here, I elaborate a simpler model first outlined in Queloz and Cueni (2021, 766).

better follow that we have reason not to come into direct contact with that to which it applies. Our concern to stay out of harm's way, our limited capacity to resist heat, or the combustible character of the physical environment we inhabit need not figure in our deliberations in terms of the concept *fire* for it to be pointful, and neither do our conceptual needs and reasons for concept use they engender. What *needs* to figure in their deliberation is the inference from the observation of smoke to the conclusion that they had better not touch whatever produces it. A parallel example, drawing on the susceptibility of human physiology to certain forms of stress, could be constructed for the more abstract concept of health (though that concept has also undergone a series of secondary elaborations that would require stringing together a series of need matrices for their appraisal). But the point is that even when our concerns, capacities, circumstances, and resulting conceptual needs do not themselves show up in our deliberations, they nevertheless explain, vindicate, and lend authority to particular ways of structuring and articulating those deliberations.

The distinction between *pro tanto* and *all-things-considered* conceptual needs can then be straightforwardly mapped onto reasons for concept use: a *pro tanto* conceptual need yields a *pro tanto* reason to use a concept. This allows us to make sense of the suggestion, in Chapter 1, that conceptual authority is a gradable and comparative notion: one might have *more* reasons, or *better* reasons, to use one concept rather than another, and one might have some reasons to use a concept that one has even stronger reasons *not* to use. Likewise, where a clear all-things-considered conceptual need emerges, it will constitute an all-things-considered reason to use a concept.

Metaconceptual reflection can therefore enable us to discern a chain of reasons running from concerns to reasons in concept use, and it will shore up certain reasons for action or belief; but our concerns are not related to our reasons for action and belief as premises to a conclusion. What our concerns yield are reasons for or against cultivating certain *dispositions*, namely the dispositions to conceptualize things in the terms specified by a given concept and to be responsive to its correlative reasons—in other words, the set of dispositions that together constitute a concept's possession conditions.¹⁵ It is only once we are disposed to conceptualize things in terms of the concept that we shall be

¹⁵ I defended the idea that reasons for concept use should be understood as being focused on possession conditions in Chapter 1, §7.

responsive to the reasons for action or reasons for belief that the concept renders articulable. Reasons for concept use do not constitute reasons from which certain reasons for action and belief *follow*; rather, they constitute reasons *to be responsive to* the reasons from which those reasons for action and belief follow.

Where the concepts are local, thick normative concepts that are not transparently indispensable in the way that concepts such as *fire* or *health* are, adducing reasons for reasons can vindicate or undermine our confidence in those concepts. If we previously harboured doubts over whether a certain part of our thought was still alive at all, or merely long-dead ballast, these reasons will dispel any lingering doubts and reaffirm our confidence. Conversely, the realization that we fail to share the concerns that would have given one reasons to use a concept can erode that concept's authority.

In vindicating or undermining our confidence in certain concepts, reasons for concept use will reinvigorate or sap the force of reasons *in* concept use. This allows us to make sense of the otherwise puzzling suggestion, which sometimes surfaces in discussions over internalism about reasons, that the reasons a concept immediately adverts to can be reasons for one agent but not for another. Williams, notably, makes that suggestion:

People who use a given concept of this sort will find their application of it guided by their experience, and also accept that it gives them reasons for or against various kinds of action. ... But this does not mean that a speaker who does use a given concept of this kind (chastity is an example that focuses the mind) can truly say that another agent who does not use the concept has a reason to avoid or pursue certain courses of action in virtue of that concept's application. To show this, the speaker would need to show that the agent has reason to use that concept, to structure his or her experience in those terms. That is a different and larger matter; all the work remains to be done. (1995d, 37–38)

While Williams is silent on what the work that remains to be done would involve, the needs-based approach offers us a detailed understanding of precisely that: by deriving a conceptual need for the concept of chastity, or something close to it, from a need matrix modelling conditions that the agent in question fulfils, one would have to show that the concept of chastity fills a conceptual need that the agent has in virtue of sharing the relevant concerns, capacities, and circumstances. In this way, a needs-based appraisal of the concept could conceivably vindicate the authority of the concept of chastity *for that agent*, and thereby show that the agent has reason to be responsive to its concomitant

reasons for action.

But equally, if it emerged that the agent failed to share the conditions that would render the concept needful, this could fatally undermine that authority. This comes out clearly in Michael Smith's explication of Williams's suggestion, which is phrased in terms that are particularly congenial to the needs-based approach:

Imagine someone who conceptualizes his experience in terms of the concept of *chastity*. Williams's idea seems to be that, for such a person, whether or not women are restraining their sexual behaviour will be a very salient feature of their circumstances. Who would have a reason to make that feature of their circumstances salient? His answer is: those who desire that women restrain their sexual behaviour. Such people have a reason to *use* the concept of *chastity*, and they have corresponding reasons for action, but all this is consistent with others who have no such desire having no such reasons. (2013, 103–4)

If we do not share the concern that renders the concept of chastity needful and see no other concern that would give us reason to use the concept, the immediate implication is that we have *no reason to use* the concept and might as well abandon it. If we were engaged users of the concept before, the concept will then be felt to have gone dead on us, and the reasons it advert to will no longer have force with us.

But if we do not *simply* fail to share the concern, the way one might simply fail to share an interest in philately, but *object* to the concern on grounds provided by concepts we continue to be confident in, we will thereby come to see a reason *not* to use the concept, because we will not want to see that concern satisfied. We will accordingly have reason to object to other people's use of that concept.

If the realization that a concept serves a certain concern is *vindictory* when the concern is one we identify with, therefore, that same realization is *incriminatory* when the concern is one we are opposed to. Just as the concerns we endorse yield reasons *for* concept use, the concerns we object to yield reasons *against* concept use. The concepts we really want to be using, on this account, are therefore the concepts that we have most reason to use, in view of our various reasons for and against concept use.

I emphasized that while there is a rational conduit leading from concerns to reasons in concept use, our concerns and conceptual needs do not typically enter our first-order deliberation as premises favouring particular courses of action. Our concerns and conceptual needs should not be understood as premises *from which* to reason, but rather

as something *in accordance with which* to reason—they are crucial parts of the extraconceptual landscape that renders certain ways of reasoning needful and pointful.¹⁶

Should we conclude from this that, when we deliberate in terms that make no reference to needs and concerns, metaconceptual reflection cannot properly enter into our deliberations and affect what reasons we take ourselves to have? Is reflection on our reasons for concept use barred from figuring in our practical deliberation on what to do in a concrete situation?

This clearly does not follow, since even deliberation in terms of first-order reasons that are not couched in terms of needs or concerns can be affected by metaconceptual reflection on our reasons for concept use if that reflection alters what concepts are brought to bear on our deliberation. In other words, there can be rational *arguments* running from concerns to reasons in concept use that directly and properly affect what we take ourselves to have reason to do in a given situation by affecting what concepts we bring to bear on the situation.¹⁷

To see how, consider first the case in which metaconceptual reflection does *not* alter what concepts one brings to bear on a situation: rather, one's reasons in concept use are shored up by reasons for concept use, thereby strengthening one's confidence in the reasons for action one saw already. The concept and its attendant reasons are provided with a bill of health. Consequently, one relaxes into thinking in those terms and being responsive to those reasons. In this case, metaconceptual reflection will not have removed first-order reasons from one's deliberation that figured there before; nor will it have introduced new first-order reasons into one's deliberation that did not figure there before. It may have revealed new *second*-order reasons to use the concept, but this will no more induce the concept to suddenly grow new reason relations than revealing the evolutionary rationale of the octopus's tentacles will induce it to grow a new tentacle. Seeing a new reason *for* the use of a concept one is already an engaged user of does not add to the reasons we see *in* the use of the concept. Mistaking the former kind of reason

¹⁶ I was led to this way of putting the contrast by Wilfrid Sellars's (1958, 286) discussion of induction as establishing principles in accordance with which we reason, but not premises from which we reason. Needs and concerns are not principles, but the general insight that one might reason *in accordance with* something though one did not reason *from* it carries over.

¹⁷ I am grateful to Sophia Dandelet and Simon Blackburn for stimulating conversations on this issue.

for a reason of the latter kind would only result in one's acting, as philosophers like to say, on the *wrong kind of reason*. All that metaconceptual reflection will have done is to give one reasons to be more confident in the first-order reasons one perceived already.

When metaconceptual reflection saps one's confidence in a concept, by contrast, one ceases to see the reasons in concept use one previously thought one had *as* reasons. This directly affects what one takes oneself to have reason to do in the concrete case, all things considered. If, before engaging in a metaconceptual reflection, one thought one had reason to feel ashamed of oneself for failing to live up to the concept of chastity, for example, then reaching the metaconceptual conclusion that one in fact has no reason whatsoever to use the concept, or even reason not to use the concept, will decisively shift the balance of first-order reasons for action away from the conclusion that one has reason to feel ashamed of oneself.

Similarly, in the case in which metaconceptual reflection reveals that one has reasons to adopt a concept one does not yet use in an engaged way, such reflection will directly affect the reasons for action one has in a particular situation. One comes to recognize *as* reasons considerations one did not previously recognize as reasons. To reverse the *chastity* example: when the young Saint Augustine, after years of living as a libertine, concluded that he had reason to adopt the concept of chastity, he came to acknowledge the force of reasons that had a direct bearing on how he had reason to conduct himself going forward. What is more, these were reasons he had reason to heed already the moment he came to recognize his reasons to use the concept; hence the amusing contradiction involved in his famous prayer: 'Grant me chastity and temperance—but not just yet'.¹⁸

Although reflection on reasons for concept use thus bears primarily on the general question of whether or not to think in certain terms, and not on the more concrete question of the reasons one has for or against particular courses of action, reasons for concept use can nonetheless end up directly affecting what reasons for action one takes oneself to have in a particular situation. Any alteration in the set of concepts we deploy in an engaged way will entrain corresponding alterations in the set of reasons for action we take ourselves to have.

¹⁸ The original reads: '*da mihi castitatem et continentiam, sed noli modo*' (1992, 8.7.17).

7.5 Answering the Authority Question

Having worked our way over the last couple of chapters from concerns via needs to reasons, we finally reach the following answer to the authority question:

The Needs-Based Conception of Conceptual Authority:

Concepts are authoritative if and to the extent that they meet the conceptual needs we have in virtue of concerns we identify with and would endorse upon critical reflection.

While this conception of authority invites us to appraise concepts by the conceptual needs they answer to, a concept's authority is not just a function of its needfulness, on this view. It is also a function of what its needfulness is rooted in, for a concept might be needful in relation to a concern we object to. Needs-based appraisal is the appraisal of concepts *together with* what creates a need for them.

On this account, we can offer the person questioning the authority of concept *F* a response along the following lines: first, guided by the expressive character of the concept, we narrow in on a concern that the person identifies with and certainly wants to see met. We then show, by constructing a need matrix or a looser discursive equivalent of it, that, given the capacities with which and the circumstances under which they pursue this concern, they have a conceptual need for something very like concept *F*. This in turn means that the person is better able to contribute to the satisfaction of the concern in question by recognizing the authority of concept *F* and responding to the reasons it advertises to than by abandoning concept *F* and trying to satisfy the concern without it. Therefore, they have a *pro tanto* reason to recognize the authority of concept *F*.

A more complex discussion ensues if our interlocutor then co-opts our approach to question that we have an all-things-considered reason to use concept *F*. Balancing the concern we highlighted against other, equally weighty concerns, they might seek to show that these concerns create *pro tanto* conceptual needs for a different concept, *F'*, and that these outweigh the *pro tanto* conceptual need for *F*, so that we would have an all-things-considered reason to use *F'*. That would of course only follow if we were forced to choose between the two concepts, and could not reasonably hope to allocate them to certain use cases in order to be able to deploy them both in coordinated fashion. But this is just the sort of metaconceptual inquiry and debate we should engage in to determine what concepts we really need.

There are, moreover, particularly complex cases that the notion of an all-things-considered reason for concept use is itself too tidy and clear-cut to handle. In those cases, the relevant nuances are better articulated in terms of a dynamic and context-sensitive array of countervailing *pro tanto* reasons. Because concerns can conflict not just between groups, but also within the breast of one individual, one can be pulled in different directions when assessing the authority of a concept. This forces one to reflect on which concerns one most identifies with, and which concerns one wants to prioritize in which contexts.

For example, even an ideological concept foisted upon me by my oppressors may be authoritative *if and to the extent that* it serves my concern to stay safe by not stepping out of line. As a result, it may be that I cannot afford to abandon it entirely. It does meet a conceptual need grounded in a concern for self-preservation.

At the same time, that very concept, if it merits being described as an *ideological* concept deployed under circumstances of *oppression*, will also radically frustrate many of my other concerns.¹⁹ I cannot then afford simply to consider it authoritative either. There are many concepts we have reason to use that we have even stronger reason *not* to use.

Furthermore, some concerns are ‘adaptive’ concerns that one identifies with, but merely as a way of adapting to non-ideal circumstances. For example, a gay man in a violently homophobic society might have a prudential concern to be perceived as a ‘real man’. Does the fact that this concern is best served by adopting the concept *real man* suffice to render that concept authoritative for him in the sense of giving him an all-things-considered reason to use it? Does he identify sufficiently with this adaptive concern to be willing to sacrifice other concerns to its pursuit, such as his concern for authenticity? How high a price is he willing to pay?

In view of the possibility of such questions, the Manichean expectation that a concept must, all things considered, either *be* or *not be* authoritative for someone across every context they face proves too simplistic. The gay man in the violently homophobic society is caught in a bind between conflicting concerns: his concern for self-preservation means that he cannot afford simply to become oblivious to the concept *real man* and the

¹⁹ For an overview of the various uses of the term ‘ideological’ and the criticism they have encountered, see Haslanger (2021). For a discussion of the methodological status of the perspective from which a concept can be characterized as ‘ideological’, see Celikates (2018).

behaviour it prescribes; at the same time, he cannot fully embrace it without betraying some of his other concerns, such as his concern for authenticity. This calls for a correspondingly complex adjustment in his cognitive economy. Through reflection on why and to what extent he has reason to recognize the authority of the concept *real man*, he may, for instance, come to acknowledge the authority of the concept only in a self-consciously prudential spirit, rendering the acknowledgement of its authority conditional on its serving his concern for self-preservation in a given context, and thereby drastically reining in the concept's influence on his life and self-conception.

Whether in the form of a decisive all-things-considered reason or in the form of a more complex admixture of contextually sensitive *pro tanto* reasons, however, reasons for concept use fundamentally answer the authority question by showing that we are better able to meet our concerns by recognizing the authority of a concept than by trying to meet them without it. Living by concepts we need really helps us to live, in the sense of furthering concerns we are identified with.

An answer to authority question along these lines of course invites the follow-up question of who 'we' is, or *whose* concerns are at issue. The use of 'we' marks the fact that, on the needs-based conception, conceptual authority is not a monolith, but a perspectival phenomenon. This perspectivalness both complicates and empowers the account.

In the first instance, the fact that different people not only have different concerns, but would still endorse different concerns after critical reflection, undoubtedly complicates the task of demonstrating the authority of a concept. It may, to a limited extent, be possible to tiptoe around that complication and vindicate some highly general concepts by drawing primarily on human beings' more constant concerns, such as the concern to avoid violent conflict with others, find out about the dangers and affordances of one's environment, secure the resources one needs to survive, and establish the conditions of cooperation.²⁰ Being maximally widely shared, such concerns promise to provide something like a starting point that people 'could not reasonably reject', in Scanlon's phrase.²¹

²⁰ See Fricker (2007, 2016b, forthcoming), Pettit (2018, forthcoming), and Smithson (2020) for accounts pursuing this sort of strategy.

²¹ See Scanlon (1998, 4 and *passim*).

But concerns like these will only take us so far. Not only does our conceptual repertoire abound with concepts that bear no relation to such constant concerns; even the concepts that can be vindicated by such concerns will still be left vastly underdetermined by them: constant concerns might show that we have reason to use something *very broadly like* concept *F*; but they will not achieve the degree of determinacy required to decide between the different conceptualizations that fall within those generous bounds.

To achieve a more determinate picture of the concepts we need, we have to leverage more *local* needs and concerns. These will give us the critical leverage and degree of determinacy required to appraise concepts in a more fine-grained way, and without restriction to anthropological universals.

However, this greater critical leverage and determinacy are bought at the cost of drawing conceptual authority from needs and concerns that are specific to one group in contrast to other groups. It follows that the concepts vindicated by local needs will be authoritative *for* that group, but not necessarily for other groups: if the authority of some concept *F* is understood in terms of its alignment with a conceptual need engendered by concern *C*, *F* will be authoritative for concept-users if and to the extent that they identify with *C* and would endorse it upon critical reflection; consequently, a given concept could be authoritative for some people and not others.

The needs-based conception thus allows for conceptual authority to be perspectival, permitting its indexation to conceptual needs and concerns that are not necessarily shared by everyone. This does not mean that there is no fact of the matter about which concepts are authoritative for whom. It is merely that the one-place predicate ‘concept *F* is *authoritative*’ turns out to correspond to a three-place relation: concept *F* is authoritative for some concept-user *U* insofar as *U* has conceptual need *N*. This is a perfectly objective three-place fact obtaining between a concept, a concept-user, and a conceptual need. Relationalism—the view that what appear to be *n*-place relations are in fact *n+k*-place relations—does not entail relativism.²² There is a fact of the matter as to what concept would best meet *U*’s conceptual needs, and this is as true if *U* is an individual person as if *U* is an entire community. To claim that concept *F* would answer to *U*’s conceptual needs is not merely to profess one’s conviction that it would be *good* if

²² See Spencer (2016).

they adopted the concept. It is to assert that it would be good *for them* if they adopted it. That is not the expression of a preference, but an empirical claim about *F*'s conduciveness to the satisfaction of certain concerns of *U*'s, given *U*'s capacities and circumstances.

The empirical nature of the claim does not, of course, release one from the requirement to *interpret* people's concerns, capacities, and circumstances so as to see them as resulting in a conceptual need. And where there is room for interpretation, there is room for deliberate misinterpretation: empirical claims about what concepts people have reason to adopt in light of their concerns can be offered in bad faith, casting the situation in tendentious terms. In practice, such abuses are often easy enough to identify under clear-headed critical scrutiny, turning, as they typically do, on flagrantly ideological assumptions about how one set of people is irremediably inferior to another set, or on brazenly self-serving accounts of what would best serve other people's concerns.

But something nevertheless needs to be said at the theoretical level to alert to, and defend against, these possibilities of abuse—even if no general defensive clause can, by itself, definitively fend off abuse, because any such clause will itself require interpretation in its application to the particular situation, and thereby lay itself open to abusive misinterpretation in turn.

A familiar refinement is the idea that the supposed improvement a concept would bring should be recognizable as an improvement to the people concerned. This qualification draws attention to the risk of distorting the situation beyond recognition, and helps defend against that risk by making the people whose concerns are at issue the ultimate arbiters of what changes to their conceptual repertoire would count as improvements.

Of course, the qualification cannot simply be taken to mean that the people in question are *in fact disposed to recognize* them as improvements, since people's actual recognitive dispositions may be faulty in various ways. The dispositions may just be a product of ignorance, or of a lack of reflection. In more complicated cases, they may be the product of uncritical exposure to distorting influences—such as systematic censorship, tireless indoctrination, and propaganda—that would lead people to move away from those dispositions if they came to understand how they had been manipulated into acquiring them.

But there is a spectrum here, and many of the contingent influences shaping one's dispositions are not so much distorting as *constructive of* our identity as a distinctive person with certain concerns in a particular time and place. Drawing the line between problematic and acceptable influences is not merely a technical matter, like sorting the wheat from the chaff; it requires ethical and political judgement on the part of the person concerned.²³ Accordingly, 'recognizable as an improvement to the people concerned' should be taken to mean: such that the people in question *would* recognize them as such if they were well informed, had given the matter plenty of thought, and continued to identify with the relevant cognitive dispositions even after understanding their formation.

It is of course tempting to go further than that, and to reach for some more objective criterion by which to ratify a set of objectively rational concerns and correspondingly objective improvements that any rational person has reason to recognize as such. To arrive at such a criterion, one would eventually be forced to look *beyond* human concerns, to some more independent standard of rightness for concepts such as the realm of Forms, the human telos, the Mind of God, natural law, the dictates of universal reason, or the structural mirror-image of the world. But it is in keeping with the humanistic anti-authoritarianism of the account of conceptual authority offered here to stop at individual concept-users' considered view of what their own concerns are and what concepts best serve those concerns. In resisting the intellectual temptation to go beyond that, the account might be thought to express a commitment to the sovereignty and autonomy of human beings. For, without supplementation by people's own judgement, it would significantly underdetermine how they should think.

Once the necessary contributions of interpretation and judgement are allowed for, however, the perspectival character of conceptual authority does not foreclose the possibility of coming to *know* that a certain concept is best *for* concept-users with certain conceptual needs. This is where, *pace* Williams, we *can* ground our confidence in thick normative concepts in a form of metaconceptual knowledge: the knowledge that a concept is right or wrong *for us*, given *our* conceptual needs.

²³ For different attempts to draw such a distinction, see Williams (2006j), Dworkin (1989), Mills (1995), Noggle (1996), Baron (2003), Greenspan (2003), Buss (2005), and Sunstein (2016). For comprehensive overviews of the relevant difficulties, see Noggle (2022) and the essays in Coons and Weber (2014).

This falls short of knowing that a concept is absolutely and definitively best; but once one really ‘goes far enough in recognizing contingency’, to use Williams’s own phrase, we can ungrudgingly recognize this, without any sense of loss.²⁴ For to realize that the standards we want our concepts to meet are as shaped by contingent forces as we are is to realize that we want our concepts to be responsive to the specificities of our situation. What we really want to *know* is whether our concepts answer to *our* conceptual needs, needs which, at most, only partially overlap with those shared by all conceivable concept-users (if any are). In many respects, our conceptual needs are distinctively ours, and the concepts that serve us best will be those that are tailored to those distinctive needs. Metaconceptual reflection on what features of our situation engender those needs, and what makes certain concepts particularly apt to meet them, can yield knowledge that certain concepts are better for us than others. And, in doing so, it can fully reconcile us to the contingency of our concepts.

On the needs-based approach, then, the thick normative concepts under which we can have some pieces of knowledge can themselves be sustained by knowledge, and not just by confidence; but this will not and *should not* be knowledge that they are absolutely and definitively the best concepts. It will and should be knowledge that they are the best concepts for us, given the conceptual needs we now have. Though this will be knowledge of a relational fact, it will itself be non-relational knowledge, and it will be as objective as knowledge in the human sciences gets. But it will not be definitive, in that our conceptual needs can and do change, so that different concepts may be called for in the future, and different pieces of metaconceptual knowledge will then have to be called upon to sustain our confidence in them.

Despite being *metaconceptual* knowledge, this will still be knowledge had *under* certain concepts, such as the psychological and sociological concepts in terms of which we make sense of our situation, our concerns, and the dynamics of the social reality

²⁴ He uses the phrase in Williams (2006g, 193), in the context of criticizing Rorty’s ironist. As my turning of Williams’s own phrase against himself suggests, there seems to me to be a genuine tension here in Williams’s work: an extrapolation of his thoughts on the authority of concepts and the significance of contingency in ‘Philosophy as a Humanistic Discipline’ puts pressure on the view he articulated in *Ethics and the Limits of Philosophy* and later reasserted in even stronger terms *after* having published his critique of ironism, namely the view that ‘[t]he thick concepts under which we can have some pieces of ethical knowledge are not themselves sustained by knowledge, but by confidence’ (1995i, 208).

around us. Needs-based concept appraisal is still a form of situated reflection, and that may render what metaconceptual knowledge it produces to varying degrees perspectival; but perspectival knowledge, as Williams himself insisted, is knowledge all the same, and when it comes to concept appraisal, it is exactly the kind of knowledge we need.

There is therefore no reason to think that a needs-based conception of conceptual authority forces us into a merely spectatorial relativism when confronting people whose conceptual needs and concerns we do not share. The concepts we are confident in continue to equip us to pass judgement on other people's concepts along with their needs and concerns, even after we recognize the authority of our own concepts to be grounded in our needs and concerns. For precisely what we shall have achieved by that recognition is the license to continue to use these concepts in an engaged way and to continue to treat them as applicable to everything they were previously applicable to, including other people's needs and concerns.

Moreover, such a needs-based answer to the authority question should not be thought of as foisting concepts upon people by *ascribing* conceptual needs to an antecedently defined set of people. Rather, the authority question gets answered in terms of reasons for concept use that are reasons for a set of people *individuated in terms of a common conceptual need*. Such an answer does not assert of a pre-specified set of people that they have a need and should therefore think a certain way; it rather asserts that *whoever* shares this conceptual need in virtue of sharing the concerns, capacities, and circumstances that engender it has a *pro tanto* reason to use this concept. In the formula: 'Concepts are authoritative if and to the extent that they meet the conceptual needs we have in virtue of concerns we identify with and would endorse upon critical reflection', the scope of the relevant 'we' is thus not independently given, but is rather a function of how far the conceptual needs at issue are shared.

The perspectivalness of this needs-based conception of conceptual authority also makes it more powerful in a number of ways. For one thing, it renders it sensitive to the specificities of people's situation. As we saw, the question should not be whether anyone has reason to use a certain concept, but whether *we* do, given *our* conceptual needs—our conceptual needs are certainly not the same as those of medieval monks or Bronze Age chieftains, and our conceptual apparatus had better be responsive to that fact.

In virtue of being attuned to local needs, this perspectival conception of authority

then allows for the possibility that, when two sets of concept-users use different concepts, that difference does not reflect an epistemic error. It might instead reflect a difference in their respective conceptual needs, itself due to an underlying difference in concerns, capacities, or circumstances.

This empowers the present account of conceptual authority to rationalize conceptual change. When people's conceptual needs change, it is only rational for their concepts to change with them. Understanding conceptual authority in needs-based terms thus makes revisions within people's conceptual apparatus intelligible as endogenous and reason-driven changes rather than as brute causal impositions by rationally contingent external factors.

By the same logic, a needs-based conception of conceptual authority can also radically change our understanding of what we are *at* when we disagree with others about which concepts to use. If we are alive not merely to the differences in our concepts, but also to the differences in our respective conceptual needs and what creates them, we become able to discriminate between situations in which we can see that others are making a mistake within a collective cognitive enterprise, as when a teacher can see that a student misunderstands the concept that is being taught, and situations in which, while we still want to say that others' concepts are *wrong*, we can at the same time recognize that they are not *simply* wrong: it makes sense *to us* that it makes sense *for them* to use the concepts they use, given how different from ours we understand their needs to be. They are not just confused, or radically deceived, or irrationally clinging to conceptual holdovers from another age. This yields the kind of understanding of where the other party is coming from that facilitates *respectful* disagreement. In Chapter 9, I will explore how much mileage we can get out of this thought in connection with pluralism in politics.

But let me end the present section by contrasting the needs-based conception of conceptual authority with other conceptions of authority in order to sharpen its contours. In its emphasis on the idea that our subordination to our concepts is conditional on *them* ultimately serving *us*, the needs-based conception of conceptual authority is a far cry from pre-Enlightenment conceptions of authority in terms of subordination and obedience to higher powers according to one's place in the Great

Chain of Being.²⁵ It is, to that extent, a humanistic, anti-authoritarian conception of authority. Instead of requiring us to conform our concepts to an order of things about which we do not have a say, it humanizes the norms to which concepts are answerable, turning these norms themselves into a function of human concerns. Some would insist that critical reflection should look *beyond* human concerns, to some more independent standard of rightness for concepts, such as the realm of Forms, human nature, the Mind of God, or the dictates of universal reason. To resist this demand is to express a humanistic commitment to the sovereignty and autonomy of human beings.²⁶ It is to opt for a picture on which we are the ones who *authorize* our concepts, and do so twice over: we not only create them, but also imbue them with authority.

At the same time, in making conceptual needs not just a function of human concerns, but also of how the world is and what capacities and limitations we bring to it, this conception of authority leaves us firmly constrained by reality in what kinds of concepts we can have reason to use. Moving away from the idea that our concepts are answerable to timeless rational foundations may invite the conclusion that we are radically free in choosing which concepts to live by; but the needs-based conception of conceptual authority suggests that our choices of concepts remain answerable to something, namely to the demands placed on us by how our concerns interact with the world in which we pursue them. We may not be forced by the world to accept the one set of concepts that is absolutely and definitively best; but as long as we bring certain concerns to the world, the world has other ways of imposing some concepts on us rather than others. Even if our thick normative concepts like *liberty* and *equality* are not answerable to the normative furniture of the universe, they remain answerable to our concerns—more precisely, to what furthering these concerns in the circumstances we face requires of us as concept-users.

²⁵ See Brandom (2002b, 2004, 2021) for a historical narrative focused on Enlightenment philosophy as a revolution in conceptions of authority. As his *Doktorvater* Rorty (2021, ch. 2) notes, anti-authoritarianism, which was one of the central characteristics of the Enlightenment, lives on notably through the pragmatist tradition, leading Rorty to think of pragmatism as an anti-authoritarianism aiming to complete the Enlightenment project by extending anti-authoritarianism from the practical to the theoretical sphere, from ethics to epistemology. McDowell (1996), Price (2011), and Brandom (2019b) can be thought of as trying to disentangle this project from the postmodernist web of confusions it became caught up in.

²⁶ See also Fricker (2020a).

In this respect, the needs-based conception does not follow Enlightenment thinkers such as Kant all the way in tying authority to rational autonomy. For those who treat autonomy as a condition on authority, nothing counts as truly authoritative over oneself unless one has freely placed oneself under its sway after critical reflection. Such autonomy-centred conceptions of authority give a central place to the will, through whose exercise alone one can bind oneself to a norm.

On the present conception of authority, by contrast, one's concepts are imbued with authority by one's conceptual needs, which in turn are a function of one's concerns, capacities, and circumstances. That is a form of authority that is neither systematically dependent on nor systematically responsive to exercises of one's will, because what conceptual needs one has is not always, or even usually, subject to one's will. If any element in that authority-bestowing structure is responsive to one's will, it is one's concerns, some of which one can simply choose to take on, redirect, or give up. But this is not true of all of one's concerns. Some of our concerns are, sometimes irremovably, part of who we are, and cannot be redirected or abandoned *ad libitum*. They are concerns directed at things we cannot help but be concerned with. And since our capacities and circumstances are, if anything, even more recalcitrant, some concepts will be authoritative for us willy-nilly. Their authority is discovered rather than freely bestowed.

And yet this conception of authority does not pit authority *against* reason, as traditionalist critics of rationalism such as Louis de Bonald or Joseph de Maistre had done in their defence of the *ancien régime*, but rather understands authority *in terms of* reason:²⁷ it is a matter of lending the prescriptions of our concepts an additional quality by augmenting them with reasons that those living by the concepts can recognize.

This conception of authority as power augmented by reasons is true to the concept's roots in Roman law: Theodor Mommsen suggests that *auctoritas*, 'authority', should be understood in terms of *augere*, 'to augment'.²⁸ Bertrand de Jouvenel also asserts that 'the root of the word denotes the idea of augmentation' (1963b, 30), and, as both he and Theodor Eschenburg emphasize, being an *auctor* to someone (*alicui auctorem esse*)

²⁷ See Friedrich (1972, 30–47) for a discussion of how these Counter-Enlightenment thinkers conceived of authority as the anti-thesis to reason and reasoning.

²⁸ See Mommsen (1888, 952, 994, 1032–39). That etymology is complicated somewhat by Heinze (1925).

meant *advising* them.²⁹ Symbolically, the adviser and augmenter was supposed to be Apollo, God of reason and moderation; in more concrete terms, however, *auctoritas* was what the Senate, composed of the *senes* (the old ones), added to the *potestas* (power) of the people.³⁰ As Carl Joachim Friedrich describes this ancient conception of authority, ‘it was a matter of adding wisdom to will, reason to force and want, that is to say, a knowledge of values shared and traditions hallowed to whatever the people wished to do’ (1972, 48). Far from constituting the non-rational counterpart to reason, on this view, authority ‘supplements a mere act of the will by adding reasons to it’ (Friedrich 1971, 19), thereby augmenting and confirming the will of the people in light of concerns that are supposed to be shared by them.

The needs-based conception of conceptual authority harks back to this idea: it likewise supplements the *power* of concepts by adding *reasons* to it, thereby augmenting and confirming the prescriptions of concepts in light of concerns shared by those who enact them.

In virtue of its emphasis on whether concepts *serve* human concerns, this conception of authority also echoes what Joseph Raz more recently described as a ‘service conception’ of authority. It conceives of conceptual authority as a matter of *serving* some concern that is a concern *for* the concept-user in such a way that the concept-user is better off living under the sway of the concept than outside it.

Yet this conception of the authority of concepts also differs in several respects from the conception that Raz elaborated over the years.³¹ His account is keyed to the issue of ‘how to understand the standing of an authoritative directive’ (2009, 134)—paradigmatically, the *state’s* directives that take the form of *law*. This focus on the directives or commands of the state leads him to concentrate on reasons of a very specific kind: reasons for action that are both *content-independent*, providing first-order reasons to do something just because one has been commanded to do it, no matter the content of the command, and *exclusionary* or *pre-emptive*, providing second-order reasons to disregard other reasons bearing on the matter.

From the perspective of the present investigation into the authority of concepts, it is

²⁹ See Jouvenel (1963b, 30) and Eschenburg (1976, 12).

³⁰ See Friedrich (1972, 129n4) and Mommsen (1888, 952, 994, 1032–39).

³¹ See Raz (1979, 1986, 1995, 2009).

clear that Raz's account only captures a special case of authority: the particularly imperious form of authority that the state exercises when it issues commands determining what *everyone* living under the state's authority really *must* do, irrespective of their inclinations and whatever other reasons they might have. These are commands that combine ubiquitous applicability with overriding stringency.

By contrast, the authority of concepts tends to be more diffuse and less imperious than the power of a state giving the law to its subjects. Concepts can be said to govern our thoughts and actions, but they do not systematically generate exclusionary and content-independent reasons for action. Most of the reasons in concept use we become sensitive to by adopting a concept lack the stringency and overriding force of commands.³² The power of concepts is a softer power.

But this does not mean that it is no real power. As we saw, once an issue is cast in certain terms, thick normative concepts can leave us nothing else to think but that a certain conclusion follows. The soft power to influence in which terms people cast an issue, and which patterns of reasoning they are disposed to follow, can be just as effective as the hard power to issue commands. If anything, the fact that it is soft power makes it *more* effective, because it enables it to be deployed surreptitiously while at the same time preserving people's sense that they are freely making up their minds—when, in a more literal sense, that is precisely what they are *not* doing insofar as they uncritically accept the make-up of the conceptual architecture they inhabit.³³

Above all, the form of authority I seek to elucidate here is prior to and underlies the form of authority that Raz elucidates. Though the soft power of concepts is different from the hard power of directives, that hard power only makes sense against the backdrop of the soft power of concepts, for the directives are stringent and exclusionary precisely

³² This also sidesteps a standard objection to Raz's account, namely that while a service conception of the state explains why the state should have the power to determine what we should do, it fails to explain why the state should also have the power to demand that we comply with its directives; see Perry (2005) and Kletzer and Renzo (2020, 205). The conception of authority I propose avoids this objection, because, while concepts help determine what we should do, they do not systematically come with a second-order demand that we comply with the reasons they generate. For further criticism of Raz's account, see Darwall (2006, 2010).

³³ See Marques (2020) as well as Queloz and Bieber (2022) for discussions of various ways in which this power can be exploited and checked.

because they need to *override* the authority of concepts. Raz conceives of authority as being *au fond* about decisions being taken out of one person's hands and put into the hands of another.³⁴ This is the authority of *authorities*. Understanding that form of authority requires understanding how directives can curtail someone's reasoning by enjoining them to disregard any reasons other than those created by the directive. The authority of concepts, by contrast, concerns the way concepts articulate and guide the reasoning that is there to be curtailed in the first place. It thus concerns a prior, more basic, and far more general form of authority, of which the authority of command-issuing authorities—when there are any—is a specific elaboration.

7.6 Wrong Kinds of Reasons

The needs-based approach thus provides us with a distinctive way of answering the authority question. But one might still worry that, by locating reasons for concept use in our conceptual needs, the needs-based approach yields the 'wrong kinds of reasons'. This worry comes in two flavours: one epistemological, the other deontological.

In its epistemological manifestation, the worry is that anything which is not an *epistemic* reason is the wrong kind of reason to motivate the use of a concept. Mona Simion, for example, writes: 'Concepts, just like beliefs, are representational devices, their function is an epistemic one: to represent the world. In virtue of this function, concepts will be properly functioning when responsive to epistemic reasons, and malfunctioning when responsive to practical reasons' (2018, 923). Consequently, she holds that 'just in the way in which prudential, moral, political, etc. considerations are the wrong kind of reasons for knowledgeable belief revision, they equally fail to support conceptual revision' (2018, 923).

But we should not be so quick to generalize from the norms governing the assessment of beliefs to the norms governing the appraisal of concepts. For one thing, these two kinds of evaluation operate at different levels. The activity of assessing or revising a belief, let us grant, is answerable only to epistemic reasons, which is to say reasons to think the belief true or false. But these are first-order reasons—reasons *in* concept use that we perceive as engaged users of the concepts in terms of which the belief is articulated. The

³⁴ See Raz (1999, 193).

activity of appraising a concept, by contrast, is answerable to second-order reasons—reasons *for* concept use that we perceive as disengaged users of the concept. Now, much as there are good second-order reasons for us to come to think of the truth as something worth pursuing for its own sake, there probably are good second-order reasons for us to only recognize *epistemic* first-order reasons as appropriate reasons for belief. But it does not automatically follow that a similar restriction holds at the second-order level of reasons for concept use. When evaluating my belief that I am currently winning some board game, for example, my concern for fairness may indeed be neither here nor there: it may seem unfair that I should be losing to the player who got to make the first move for the fifth time in a row, but that makes it no less true. When the designers of that game defined the concepts governing its dynamics, however, they quite properly let their concern for fairness inform their conceptual choices—at that level, the thought that some conceptualization would systematically offend against fairness is clearly the right kind of reason.

The wrong-kinds-of-reasons worry is further alleviated if we do not think of all concepts as being exclusively in the business of turning the intellect into a mirror of the world, but as performing numerous other roles besides. Tellingly, Simion's guiding example in generalizing from belief revision to concept appraisal is what she takes to be a joint-carving natural kind concept: 'If our concept of "deer" is epistemically perfectly functional and carves nature at its biological joints', she writes, 'moral, political, etc. considerations, in isolation, will not be the right kinds of reasons to revise it' (2018, 923). If the idea that our concepts *exclusively* serve the epistemic function of representing the world has any plausibility, it is in connection with natural kind concepts such as *deer*.

But while, from an engaged perspective, even thick normative concepts seem to be in the business of mirroring the antecedent structure of social reality, that impression dissipates once we consider them from a disengaged perspective and pay attention to the varieties of work they actually perform: many of our moral, political, or legal concepts have been thought to perform a host of different functions which, from a disengaged perspective, have nothing to do with limning worldly boundaries, and more to do with shaping and regulating society and expressing norms of various kinds.³⁵ This functional

³⁵ See Blackburn (1993, 2013a, b, 2017), Brandom (1994, 2001, 2011, 2013, 2015b), Price (2011; 2013; 2007), Williams (2010, 2013), and Thomasson (2015, 2020a, b, 2022).

pluralism should make us hesitant to generalize from a natural kind concept like *deer* to moral, political, or legal concepts, or indeed to logical, mathematical, and modal ones.

Moreover, notice that even a concept like *deer* does not cater exclusively to the disinterested concerns of taxonomists. It is a concept that simultaneously needs to tie in with a host of other human concerns, ranging from the predatory and the culinary to the cultural, symbolical, mythical, and theological. It is not least the pull of these other, not purely epistemic concerns that explains why taxonomies themselves tend to have tortuous histories, and why, to this day, '[c]ontemporary biology seems committed to pluralism, as different investigators use the classifications best suited to their needs' (Kitcher 2001). Even the prototypical example of a natural kind concept, the concept of water, has been thought to answer to a panoply of cross-cutting concerns due to its longstanding enmeshment in everything from agriculture and transport to cooking and religious rituals.³⁶

When shifting our focus away from natural kind concepts towards the thick normative concepts that organize social worlds, the prospect of understanding conceptual authority exclusively in epistemic terms and the function of concepts exclusively in representational terms becomes even less inviting. Thick normative concepts do not even seem to play purely epistemic roles: they manifestly derive their importance from their social roles in forming conceptual architectures that enable coordination and cooperation. As Tadeusz Zawidzki remarks, '[o]ur social accomplishments are not by-products of individualized cognitive feats' (2013, xiii). One cannot satisfy the conditions of coordination and cooperation through epistemic prowess alone. It is the collective cultivation of thick normative concepts that shapes our minds for cohabitation in society. Such mindshaping is not an epistemic feat, but a practical one.³⁷ In the first instance, thick normative concepts do not help us to more accurately mirror antecedent objects and properties, but rather help us to create a social world by instituting norms and practices, establishing institutions, and regulating communal life.³⁸ It therefore seems only right

³⁶ See Schroeter and Schroeter (2015, 426).

³⁷ The term 'mindshaping' was introduced by Mameli (2001) and taken up by Zawidzki (2013).

³⁸ Sally Haslanger urges a similar shift in perspective: 'we need to bring to the forefront the role of language in shaping us for coordination. ... Especially in the case of social kinds, the evaluation of a linguistic/conceptual tradition should consider how it shapes us to be responsive to each other (and the

that our confidence in concepts performing such moral and political roles should be responsive to moral and political reasons. Here, to insist that any reasons for concept use that are not epistemic must be the wrong kinds of reasons would look less like a compelling objection than like an epistemologist's *déformation professionnelle*.

What is more, the reasons that *are* epistemic anyway require supplementation by other kinds of reasons in this connection, because they cannot take us far enough by themselves. One can argue that some concepts are worth having because they enable us to articulate knowledge. But different concepts would enable us to articulate different forms of knowledge. So while it may be true that certain concepts put certain forms of knowledge within our reach, this fact alone is not enough to make a neutral case to prefer these concepts over equally knowledge-generating alternatives. It merely leads on to the question of what makes certain kinds of knowledge *more worthwhile than others for us*. And this question calls for non-epistemic reasons adducing facts about us—about our concerns, capacities, and circumstances, and the needs for certain kinds of knowledge we have as a result.

But perhaps the more familiar manifestation of the wrong-kinds-of-reasons worry is its deontological incarnation: the worry that anything that is not a *moral* reason must be the wrong kind of reason to motivate the use of a concept. While the needs-based approach may, for instance, indicate various prudential reasons to recognize the authority of a concept such as *moral rightness*, the worry goes, moral rightness is precisely not authoritative merely prudentially—to do the morally right thing merely out of prudence is to do it for the wrong kind of reason.

The first point to emphasize in response is that we must resist the slide from a concept to its object: confronted with the claim that we have reason to use the concept of x because it is instrumental to meeting some conceptual need, philosophers tend to be quick to point out that the value of x is not merely instrumental, and that to pursue x because it serves some ulterior concern is to act on the wrong kind of reason. But the value of x is one thing, the value of the *concept* of x quite another. Knowledge as a mental state, for example, can be valuable in various ways, both instrumentally and intrinsically. But the value of the *concept* of knowledge is a distinct issue. Its value is not the value of a

world), what forms of social reality are created (institutions, practices, artifacts, identities), and whether or how we might do better' (2020a, 249).

state, but the value of a cognitive device that notably allows us to *recognize* knowledge as such and to think *about* knowledge. Were we not a social and language-using species that shares information, we would not have the particular conceptual need for something like the concept of knowledge that we highlighted in the previous chapter. But we would still need knowledge itself, especially concerning our immediate environment and its threats and opportunities. It can therefore be granted that the wrong-kinds-of-reasons worry should be taken seriously in thinking about the *objects* of our concepts. But this does not preclude our concepts from standing in instrumental relations to our concerns, even if the view we take of things when thinking in those terms is not an instrumental view, but one on which they possess intrinsic value, or are things that simply *must* be done. Instrumentality does not entail instrumental mentality.

Another important point is that the concerns in relation to which concepts are instrumental can be the most high-minded *moral* concerns, and when they are, the resulting reasons for concept use will be *moral* reasons—we might have reason to jettison one concept in favour of another out of a concern for fairness, or justice, or impartiality. Nothing restricts the needs-based approach to reasons of a prudential kind. If we have reason to use a certain concept of moral rightness to realize our moral concerns, that makes neither the reasons nor the concerns prudential.

Part of what fuels the wrong-kinds-of-reasons worry is the way in which kinds of reasons themselves are conceptualized. ‘What misleads people’, Harry Frankfurt observes, ‘may be the supposition that the only alternative to accepting the requirements of morality consists in greedily permitting oneself to be driven by self-interest’ (2004, 8). If we think in terms of a such a *dualistic conception of reasons*, all reasons for action must *either* be reasons of morality *or* reasons of self-interest. Accordingly, whatever is not a moral reason must be a merely prudential reason.

But there are other kinds of value and importance besides the moral and the prudential. The mere fact that a reason is not clearly moral does not mean that it should be relegated to the rank of the merely prudential. Most of the reasons that make the world go round lie between the extremes that this overly stark contrast presents as the only options. As Susan Wolf has vividly brought out, most people find their reasons to live not in reasons of morality or reasons of self-interest, but in ‘reasons of love’, engendered by

their love of persons they are attached to or pursuits they are passionate about.³⁹ These reasons that the dualistic conception omits ‘are some of the most important and central ones in our lives’, which ‘engage us in the activities that make our lives worth living’, ‘give us a reason to go on’ and ‘give meaning to our lives’ (2010, 2). As long as philosophers remain hostage to the dualistic conception of reasons, they risk ignoring much of what actually moves people—including, ironically, what moves them to do philosophy.

If reasons for concept use are not to be prudential reasons, on this dualistic conception, they would have to be moral reasons. But when it is not an action or a motive but a *concept* that is in the dock, morality’s evaluative machinery, keyed as it is to the appraisal of individual actions and the first-order reasons for which they were performed, can make it hard to see how reasons that float at one remove from action, such as reasons for concept use, could be *bona fide* moral reasons—especially if they do not take the form of moral obligations. The dualistic conception of reasons then conspires with the focus of many moral concepts on actions and their first-order motivations to produce the impression that all reasons for concept use must be merely prudential.

Three observations help dispel that impression, however. First, morality’s evaluative machinery *can* be brought to bear on concepts, and even when it is understood along Kantian lines, it plausibly makes at least *some* demands on one’s conceptual apparatus. For example, if morality demands not only that one fulfil one’s moral obligations, but that one fulfil them *because* they are one’s moral obligations, this already imposes demands on one’s thought as well as on one’s actions: it places one under a moral obligation to use whatever concepts are required to be able to act from the right kind of motive. Depending on how exactly the demand to do the right thing from the right kind of motive is understood, there may be several ways in which one’s conceptual apparatus can satisfy that demand. It may simply do so by putting at one’s disposal the concept of moral obligation or moral duty itself, thereby enabling one to φ from the thought ‘Because I am under a moral obligation to φ ’. But equally, it may satisfy the demand by putting at one’s disposal a special concept of *ought* that expresses moral obligation, so that one can φ from the thought ‘Because I ought to φ ’. Or, if the demand is understood to allow for this, one’s conceptual apparatus may even satisfy it by equipping one to register one’s various moral obligations under a host of more particularized and concretized descriptions: when

³⁹ See Wolf (2010, 5–6); Frankfurt (2004) also refers to these underappreciated reasons as reasons of love.

under a moral obligation to keep a promise, for instance, the description under which that obligation forms the motive from which one acts might take the particularized form ‘Because I promised’.⁴⁰ In whichever way the demand is met, however, there is a set of concepts such that one is under a moral obligation to use at least one suitable subset of them.

Second, some of the concerns that concepts serve will be moral concerns, and the reasons for concept use engendered thereby will be moral reasons, even if the relations between concepts and concerns are instrumental. The mere fact that the use of a concept is instrumental to realizing a concern does not make the concern itself instrumental. One’s concern for equality, say, might give one reason to adopt concepts whose use promotes equality, and to abandon concepts that obstruct or frustrate that concern. Furthermore, this need not involve demoting concepts to the status of mere means, devoid of anything but instrumental value. In some cases, the use of a concept might itself *instantiate* the realization of the concern it serves. As we saw Murdoch argue, a change in concepts might be not just instrumental to, but constitutive of moral progress.⁴¹

Third, even the concerns that are *not* moral concerns need not therefore be self-interested or prudential. Like our reasons for action, reasons for concept use should not be conceived as being exhausted by reasons of morality and reasons of self-interest. Many of our most forceful reasons for concept use might be second-order reasons of love, reflecting conceptual needs engendered by our subjective attachment to objectively worthy pursuits. Philosophers should know, since it is surely their love of philosophy as much as their sense of duty or self-interest that leads them to so tirelessly analyse, refine, replace, reject, rehabilitate, and create concepts. Their philosophical passions and projects give philosophers reasons to use the most abstruse concepts that no one else has reason to use, and that neither morality nor self-interest are well served by. The same reasons which, on Wolf’s account, give meaning to our lives can also give authority to our concepts.

⁴⁰ Williams (1981d, 117) holds that the demand is more charitably interpreted that way.

⁴¹ See Murdoch (1956, 1961).

7.7 Conceptual Ethics vs. Conceptual Morality

In some cases, of course, the reasons we might have to use a concept will indeed include genuinely prudential reasons, and here, deeper differences in philosophical outlook will emerge in whether one recognizes these reasons as pertinent, or dismisses them as the wrong kind of reason. Echoing Williams's distinction between ethics and morality, we might mark these differences by conceiving of *conceptual ethics* as the broad-minded enterprise of appraising concepts on the basis of *all kinds* of considerations that can inform how one should think and live, and contrast this with the more narrowly focused enterprise of *conceptual morality*, understood as a purely moral evaluation of concepts that insists on drawing a sharp boundary between moral and non-moral considerations and only counts moral considerations as pertinent.⁴²

McPherson and Plunkett, for example, might be taken to abstractly express the exclusive focus of conceptual morality on a specific kind of normative consideration when they write: 'what is crucial is whether use of a normative concept is vindicated by *specific* normative concepts: namely, the most authoritatively normative concepts' (2021, 218). We should rely on normative concepts such as *morally right* or *morally better* in appraising other concepts, on their view, but not on the wider repertoire of thick normative concepts that we also have at our disposal. Accordingly, if we engage in conceptual morality as opposed to conceptual ethics, we will be led to appraise all our concepts in terms of a narrow set of moral considerations, such as their tendency to maximize some general currency like well-being, or their alignment with our moral obligations. An example of the former would be indirect utilitarianism about concepts. An example of the latter would be Christine Korsgaard's preferred Kantian way of demonstrating 'the right of ... concepts to give laws to us' (1996, 9), which ultimately also amounts to a purely moral evaluation that insists on drawing a sharp boundary between moral and non-moral considerations and only counts moral considerations as pertinent, dismissing other types of considerations as the wrong kinds of reasons.

Even if one were to accept that the *reasons for action* bearing on one's practical deliberation on what to do should be limited to moral reasons taking the form of moral obligations, however, it still would not automatically follow that this restriction to

⁴² See Williams (2011, 7).

moral reasons extended to the *reasons for concept use* bearing on the question of what concepts to recognize as authoritative. More philosophical work would be required to show, first, that the question of conceptual authority is even a *practical* question in the relevant Kantian sense, which is to say a question as to what one ought to do, where the ‘ought’ expresses moral obligation; and second, that it is even intelligible to speak of ‘moral obligations’ to use or abstain from using a concept *for any kind of concept*, and not just in exceptional cases, such as when it comes to the concept of moral obligation itself.

By contrast, the needs-based approach, in its willingness to recognize even prudential concerns as legitimate normative input to the authority question, can be characterized more closely as exemplifying conceptual ethics *as opposed to* conceptual morality. The needs-based approach advocates a thoroughly pluralistic appraisal of concepts, which in principle allows us to draw on any among the rest of our concepts, including thick normative concepts as well as prudential, political, or aesthetic ones. In inviting us to ask what concepts we need, it therefore imposes no principled restriction on what we might need them *for*—everything we care about is admissible, if defeasible, input to concept appraisal: all our various concerns are potentially relevant, and not merely, say, the concern to maximize well-being. In Chapter 9, for example, we shall consider an illustration of needs-based concept appraisal that appeals to a distinctly nonmoralistic admixture of prudential and political considerations.

7.8 Conceptual Good-for-Nothings

Locating the value of concepts in their aptness for serving our concerns by meeting our conceptual needs might ultimately still seem reductive in at least one respect, however, namely insofar as it appears to reduce the goodness of concepts to what they are good *for*. And it is equally a theme of Wolf’s work that cultural artefacts need not always be good in virtue of being good *for* something—in particular, they need not be good in virtue of making some measurable contribution to well-being.⁴³ Some artefacts are valuable even

⁴³ An observation echoed by Scanlon (1998, 143). Insofar as contributing to our well-being is a narrower idea than serving our concerns, because not all our concerns are directed to our well-being, I am broadening the contrast class here.

though they are *good-for-nothings*. One's first acquaintance with 'a poem or a novel or a painting', for example, can have 'the character of a discovery of something valuable in itself' (Wolf 2015a, 76).

To make sense of such valuable good-for-nothings, Wolf invites us to consider the view that

realizing our intellectual and perceptual potentials is good-in-itself. ... [A] part of human good involves being connected in appropriate ways to what the world has to offer. ... [I]f we understand the world as containing objects and opportunities for experience that are of value in themselves, then we may think of our lives as better, as more fortunate, insofar as we are able to be in appreciative touch with some of the most valuable of these. (2015a, 76)

Wolf primarily has works of art, philosophy, and science in mind; but concepts are cultural artefacts too, and sometimes form the backbones of innovative works of art, philosophy, or science. Can concepts also be valuable good-for-nothings? Might this be the mite of truth that the wrong-kinds-of-reasons worry points us towards?

Wolf remarks that what makes philosophy good are things like 'illuminating a problem' or offering novel ways of 'interpreting our experience' and 'understanding our relation to the world' (2015a, 85). Yet these are the same sorts of things that concepts achieve. And sometimes, acquiring a new concept does have 'the character of a discovery of something valuable in itself' (2015a, 76). The brilliant distinction, the delightful nuance, the satisfaction of discovering a different way of seeing things—these will be recognizable to philosophers as experiences of value.

Perhaps, therefore, we can intelligibly value a concept simply for what it allows us to think, just as we can value a certain food for its taste rather than for its nutritional benefits. Whatever else it does, a new concept allows us to explore previously inaccessible aspects and dimensions of the world. And when several new concepts band together, they can help us overcome the constraints of established ways of thinking, sharpen our senses, and open up new forms of knowledge.

If, as Wolf contends, 'realizing our intellectual and perceptual potentials is good-in-itself', this suggests that we should strive for a richer conceptual repertoire, even if the concepts are good-for-nothings. Indeed, Wolf herself introduces a concept that articulates a highly general motivation for enriching our conceptual repertoire: *love of the world*—'an attitude in which life seems endlessly fascinating, yielding countless objects

of interest and admiration' (2015b, 177). Love of the world can drive the proliferation of concepts and distinctions as an end in itself.

But does the conceptual proliferation encouraged by this attitude have a sense of direction? Or should we see value simply in the multiplication of concepts and distinctions? Surely, the idea cannot simply be that *more* concepts and distinctions is always *better*. Even Williams, for whom '[o]ur major problem now' is that we have 'too few' (2011, 130) ethical concepts, warns against chasing 'the shudder of an exquisite distinction' (2005g, 53) just for its own sake. The enrichment of our conceptual repertoire, even if intrinsically valuable, still requires guidance by a sense of quality. We must remain able to discriminate between enrichment and encumbrance, between careful nuancing and frivolous hair-splitting. Thus, when Murdoch argues that we should acquire more concepts in terms of which to picture the substance of our being, because some moral improvements can be achieved already by changing how we perceive a situation, for example, she still invites the more discriminating question of what makes certain concepts progressive, and what makes a particular way of seeing the situation important.⁴⁴ We cannot do without some guiding sense of what forms of conceptual diversity are worth having.

To put the problem of undirected conceptual proliferation back in its box, we need only remember that we are already well equipped to discriminate between pointless hair-splitting and worthwhile distinctions—but not because worthwhile distinctions are always identifiable antecedently of human beings and their concerns. It is rather *in virtue* of the concerns and concepts we bring to the appraisal of concepts that we can discern value in certain distinctions, or importance in certain ways of seeing a situation. This shines through in the way Wolf herself invites us to see value in certain good-for-nothings: she still appeals to human concerns that are furthered, instantiated, or realized by these good-for-nothings, such as the concern *to realize our intellectual and perceptual potentials*, or the concern *to appreciate what the world has to offer*. These are highly general epistemic and aesthetic concerns that may indeed be nothing like Wolf's contrast foil, the welfare theorist's hard-headed concern to deliver measurable benefits to longevity and health; but they are among the most characteristic of human concerns.

The fact that Wolf explains the value of good-for-nothings by tying them back to

⁴⁴ See Murdoch (1956).

human concerns is not just an incidental feature of her examples, but reflects a broader hermeneutic constraint on methodologically humanistic reflection: if, as Wolf herself declares, we want no metaethical commitment ‘to a Platonic world of ideas and values that are independent of human existence’ (2015a, 77), the value of concepts must be related to the human point of view at least at the level of reflective explanation. This does not entail that we can only value concepts narrowly in terms of their instrumentality in meeting human concerns. But it does entail that our seeing inherent value in concepts expresses a human attitude, and that we must be able to make reflective sense of that attitude in humanistic terms, by seeing how it meshes with the rest of human affairs. The constraint is therefore this: to be able to make reflective sense of how certain concepts can be valuable in their own right, we need to be able to see how the attitude of valuing them in this way relates to some recognizable human concern. As long as the attitude of valuing certain concepts for their own sake remains disconnected from *any* human concern, the claim that those concepts are inherently valuable will not be fully intelligible to us under reflection.

It follows that even the goodness of conceptual good-for-nothings cannot ultimately be completely independent of human concerns. If we entirely lacked certain concerns, the delightful nuance, the exciting concept, the important difference in how we see a situation—these would be stripped of their significance. Their inherent value depends on their enmeshment in a certain practice animated by certain concerns.

Insisting that even conceptual good-for-nothings must still tie in with some human concern at this reflective level puts the problem of undirected conceptual proliferation back in its box, since the extension of our conceptual repertoire can draw guidance from its roots in our concerns. By relying on the concerns and concepts we possess already, we can discriminate between worthwhile distinctions and pointless hair-splitting.

At the same time, once we replace the dualistic conception of reasons with a more pluralistic picture of what moves people and reflect on how even seemingly idle concepts might fill conceptual needs after all, this puts pressure on the idea that conceptual good-for-nothings really are good for *nothing*. If only at the reflective level of philosophical explanation, the concepts we experience as inherently valuable are in fact good for *something*: realizing one’s intellectual and perceptual potentials, notably, and appreciating what the world has to offer. It is only on a restricted understanding of ‘being good for

'something', such as the welfarist conception of goodness, that concepts are intelligible as good-for-nothings at the reflective level.

Even so, Wolf's illumination of the value of good-for-nothings helps us appreciate that we value some things, including some concepts, for their own sake. What I have sought to add is that even these forms of inherent value must retain some connection to recognizable human concerns at the level of philosophical reflection if they are to make sense to us in humanistic terms. This addition achieves two things: it indicates on what basis we might discriminate between more or less important additions to the conceptual repertoire; and it reinforces the point that we need a more nuanced picture of human reasons and motives to appreciate the range of ways in which things, including concepts, can be valuable.

Thinking through the value of conceptual good-for-nothings thus leads us back to the importance of reasons of love as reasons for concept use. If there is value in acquiring new concepts, appreciating fresh nuances, and becoming sensitive to finer distinctions, it is because we have more reasons for concept use than the dualistic conception would have us believe. We are concerned not merely to increase our own well-being or that of others, but to realize our intellectual potentials, be attentive to the world around us, and appreciate what it has to offer. With her notion of *love of the world*, Wolf fashions a valuable conceptual lens that renders these invisible concerns visible—one philosophers have every reason to use.

This concludes the presentation of the needs-based approach to concept appraisal. In the next chapter, we turn to a case study of how this approach can cast fresh light on a contested concept at the centre of debates over free will and responsibility: the concept of doing something *voluntarily*. Working through this case study will not only further illustrate the needs-based approach, but also underscore two significant insights it yields: that sometimes, powerful concerns can distort concepts out of the shape in which they best serve the balance of our concerns; and that sometimes, there are good reasons for us to favour concepts exhibiting what the tidy-minded view considers a defect, namely superficiality.

CHAPTER EIGHT

The Essential Superficiality of the Voluntary

8.1 A Questionable Concept

What makes an action voluntary? And does it matter? Is the concept of the voluntary even an important one to have in our repertoire?

Recent history has not been kind to the concept of the voluntary. For centuries, this concept, which marks out actions attributable to an agent's *voluntas* or will, had been at the heart of debates over whether we have free will, how the mind directs the body, and which actions we should be held responsible for. It had formed a centrepiece of accounts of action from Descartes through Locke, Berkeley, Hume, Reid, and Bentham to Mill.¹ They all felt the urge to *deepen* the concept by theorizing it in terms of a detailed metaphysical account of the causal underpinnings that make a bodily movement voluntary. What is more, they all broadly agreed that what turns a bodily movement into a voluntary action is its being caused by a *volition* or an *act of will*.² Exponents of this 'theory of volitions' also included Thomas Brown, whose influence on the nineteenth-century legal theorist John Austin ensured the theory's lasting impact on Anglophone jurisprudence.³

¹ For a historical overview of the development of the theory of action, see Hyman (2011; 2015, 1–24). See also Wilson and Shpall (2012), Candlish and Damnjanovic (2013), D'Oro and Sandis (2013), and Glock (2014).

² Though there were also dissenters, such as Thomas Hobbes and Alexander Bain.

³ See Brown (2012, Part I, section 3) and Austin (1885, Lecture XVIII, 411–15).

These attempts to deepen the concept by understanding it in terms of its detailed causal underpinnings implicitly endorsed the idea that greater conceptual authority could be achieved by realizing theoretical virtues such as depth. But the most obvious attraction of a deepened concept of the voluntary lay in its promise to give philosophers an objective, independent yardstick by which to take the measure of our practices of responsibility attribution. By first developing a self-standing account of what, at the deepest causal level, *makes* an action voluntary, we achieve a metaphysically grounded understanding of which actions, if any, we are truly responsible for; and given such an independent understanding of responsibility, we can then take the measure of our actual practices of holding people accountable for their actions: we can determine to what extent *attribution*s of responsibility can be grounded in *true* responsibility.

Yet the twentieth century saw the theory of volitions collapse under three successive waves of criticism. William James and Bertrand Russell maintained that postulating volitions was not required to make sense of action: action arose when the memories of kinaesthetic sensations first experienced in mere bodily movements were recruited to function as ‘motive ideas’.⁴ Ludwig Wittgenstein and Gilbert Ryle then argued that while the distinction between voluntary and involuntary actions was innocuous enough, we should resist the view that there was, as Wittgenstein put it, ‘one common difference between so-called voluntary acts and involuntary ones, viz, the presence or absence of one element, the “act of volition”’ (1958, 151–2). In fact, such a view engendered a regress, as Ryle pointed out, for what about the act of will itself? If it was itself voluntary, one had to explain its voluntariness in terms of another act of will, and that act’s voluntariness in terms of yet another act of will, and so on, *ad infinitum*; if it was not voluntary, on the other hand, this had the equally uninviting consequence that the source of an action’s voluntariness would have to lie in an involuntary act of will.⁵ Finally, G. E. M. Anscombe and Donald Davidson, who each ushered in what are still the reigning paradigms in contemporary thought about action, both influentially insisted that the best entry-point for action theory was not the concept of the voluntary, but the concept of the

⁴ See James (1981, ch. XXVI) and Russell (1921, 285)

⁵ See Ryle (2009a, 54). For rejoinders to Ryle’s criticism in particular, see Hornsby (1980, 48–50) and O’Shaughnessy (2008b, 363–84), and see Alvarez and Hyman (2019) for an overview of the development of the theory of action in the second half of the twentieth century.

intentional.⁶

Since then, as John Hyman observes, the concept of the voluntary has largely been neglected. ‘The theory of volitions had been demolished’, he explains, ‘and the delicate task of lifting voluntariness out of the ruins did not seem worth the trouble’ (2015, 75). It is true that the concept continues to figure prominently in criminal law,⁷ and, as we shall see, some form of it also underpins sceptical challenges to free will. But, in the theory of action, the concept came to look increasingly off-putting as the time-honoured metaphysical constructs attempting to explain what makes a bodily movement voluntary fell into disrepair.

In view of this inauspicious history, there is a real question over whether we still have any reason to use the concept of the voluntary, or whether its authority has been terminally eroded by the failures of metaphysical accounts of voluntariness.

This is a question in conceptual ethics that the needs-based approach is ideally poised to answer. For, instead of trying to derive the contours of the concept of the voluntary from a metaphysical inquiry into the nature of its extension, the needs-based approach pursues the reverse methodological strategy, proposing to let our understanding of what kind of concept of voluntariness (if any) we have most reason to use grow out of an understanding of our concerns and conceptual needs. In other words, it aims to work from the concerns animating the use of the voluntary/involuntary distinction to those versions of the distinction that are worth using. Should we use a concept such that an action’s being causally determined already forecloses its counting as voluntary? Or should the concept of voluntariness be properly applicable to actions that are causally determined but uncoerced? What about uncoerced actions performed in an unusual state of mind—in the grip of extreme passion, for example? Turning to what lies downstream of the concept’s application, what inferential consequences should the concept’s applicability have? How should the concept tie in with moral and legal practices

⁶ See Anscombe (1957) and Davidson (1980); though it must be said that Anscombe’s views on the concept of the voluntary and its ethical role in her 1960s essays and in her second McGivney lecture (2008b) differ notably from her remarks in *Intention* (1957, §§7–17, 20–22, 49). She even considered variations on the concept that would be applicable to non-human animals if we spoke of ‘desire’ instead of ‘will’; see Bierson and Schwenkler (2022, 329n40).

⁷ See, e.g., Moore (2010, 5) and Saunders (1988).

of holding people responsible for some actions and excusing them for others?

What the needs-based approach suggests is that this battery of questions can be answered by looking at the conceptual needs to which the concept of the voluntary is answerable. Our concerns, when combined with our capacities and circumstances, can provide reasons for us to use the concept of the voluntary in *some* form; and, once our understanding of the various needs to which it answers becomes fine-grained enough, those needs can even provide reasons for us to use *certain* forms of the concept rather than others.

Applying this approach to the concept of the voluntary will also offer a further illustration of how the needs-based approach has the edge over the tidy-minded approach through its ability to register the practical virtues of theoretical vices. In particular, it will bring out the value of the *superficiality* of the concept of the voluntary.

The importance of the concept's superficiality has been overlooked by advocates and critics alike. Theorists who deemed the concept of the voluntary important have tended to do so because they considered it an enigmatic but profound idea that could be deepened through a theory of action, while those who deemed it a superficial idea that could not coherently be deepened have tended to neglect it as unimportant.

Parting company with both camps, I contend that the needs-based approach reveals the concept of the voluntary to be at once important *and* superficial—indeed, to be important *only as long as* it remains superficial. The concept is an *essentially* superficial concept that performs important work for us, but it can only perform that work if, in contravention of the tidy-minded view of superficiality as a defect, we refrain from deepening the concept. Any *viable* version of the concept—i.e. any version capable of serving our concerns effectively—will be superficial, because the concept cannot serve our concerns if it is deepened. In saying that the idea of the voluntary is ‘essentially’ superficial, I am therefore not suggesting that the idea has some discoverable ‘essence’ in light of which it turns out to be superficial—precisely not; the argument runs the other way, from the concerns fuelling our interest in the voluntary/involuntary distinction to those versions of the distinction that are worth having.

Something like this view is adumbrated in Anscombe’s essays from the 1960s,⁸ but it

⁸ See Anscombe (2005, 2008a). For an account of how Anscombe’s views on the voluntary and its ethical role change after *Intention*, see Bierson and Schwenkler (2022).

is most explicitly articulated in the following passage from Bernard Williams's *Shame and Necessity*:

The idea of the voluntary ... is essentially superficial. It is a mistake to suppose that the notion of the voluntary is a profound conception that is threatened only by some opposing and profound theory about the universe (in particular, to the effect that determinism is true). That supposition underlies the traditional metaphysical problem of the freedom of the will. ... Just as there is a 'problem of evil' only for those who expect the world to be good, there is a problem of free will only for those who think that the notion of the voluntary can be metaphysically deepened. In truth, though it may be extended or contracted in various ways, it can hardly be deepened at all. What threatens it is the attempt to make it profound, and the effect of trying to deepen it is to put it beyond all recognition. (1993, 68–69)

Though Williams may sound like a complacent compatibilist in this passage, we shall see towards the end of this chapter that this impression is misleading.⁹ The question more immediately raised by this passage, however, is how exactly a superficial concept of the voluntary contrasts with a deepened one, and what the thesis of the 'essential superficiality' of that concept amounts to.

I propose to develop this thesis as an illustration of how the pursuit of conceptual authority through depth can be misguided. In doing so, I shall consult Williams's scattered remarks on the superficiality of voluntariness.¹⁰ But I shall have to go substantially beyond Williams in doing so, since he did not develop the suggestion himself. Nor was it developed in the direction I envisage by the subsequent literature.¹¹ Williams's suggestion is alluded to, but not really discussed, in Duff and von Hirsch

⁹ As Paul Russell also concludes after examining Williams's (1995c, 6) criticism of the 'reconcilers' and the 'old compatibilism': 'Whatever final position Williams arrives at, it should not be understood as any form of comfortable or complacent compatibilism' (2022, 178). On P. F. Strawson's (2008a) influential framing, Williams sides neither with 'the pessimist' nor with 'the optimist', but rather indicates a third way incorporating insights from both, thereby arguably adopting the sort of position Strawson himself advocated: see Russell (2017c), Queloz (2021c), De Mesel (2021), and Emilsson (manuscript).

¹⁰ See Williams (1993, 67; 1995a, 578; 1995b, 127–28; 1995e, 243, 247n5; 1995n, 495; 2006e, 124–25).

¹¹ Though I have benefited from Moore's (2003, 2006a) and Louden's (2007) reconstructions of Williams's critique of the morality system as targeting the concept of a 'purely voluntary act', in Moore's apt phrase. See also Queloz (2022b) as well as Krishnan and Queloz (2022) for discussions of that aspect of Williams's critique.

(1997, 103), Matravers (2007, 53, 57), and Crisp (2017, 1), and insofar as it has been unpacked at all, it has been interpreted as a dismissive remark pointing to superficiality as a defect, with the qualification ‘essentially’ meaning only ‘basically’ or ‘at bottom’: Yeager (2006, ch. 2), for example, takes the superficiality of the idea of the voluntary to be a flaw to be remedied, while Deigh (2008, xi) takes it as an encouragement to move away from the idea of the voluntary in our practices of moral appraisal.

By contrast, I take the superficiality of the concept of the voluntary to be neither a flaw nor a reason to rely less on it in our practices of moral appraisal. On the contrary: the concept is very much worth having, and its superficiality is not a defect, but an important feature of it, one that is ‘essential’ in that the concept *cannot function properly without it*. We not only need the concept of the voluntary, but we also need it to be superficial. A great deal turns on the issue, moreover, because the concept marks a key point at which our psychological concepts link up with our moral concepts, including notably our concepts of responsibility. Deepening the concept of the voluntary thus threatens to rob attributions of moral and legal responsibility of their efficacy in helping us to live together.

In speaking of ‘the’ idea of the voluntary, I do not mean to deny that there are several different, though related, concepts that we express with the word ‘voluntary’.¹² The concept that the word expresses in its moral use, for example, may not be exactly co-extensive with the concept it expresses in its legal use; moreover, the boundary between voluntary and involuntary action may be blurred in the word’s moral use, allowing for actions that lie halfway between the fully voluntary and the utterly involuntary, but sharp in at least some of its legal uses: in criminal law, for instance, an accusation must issue in a verdict of guilty or not guilty, and this may force the binary classification of actions into either voluntary or involuntary ones, even if some actions are neither clearly one nor clearly the other.¹³ A sufficiently close look at ‘the’ concept of the voluntary may thus find that, in different contexts, it differentiates into a collection of related but subtly different

¹² Anscombe’s own use of the term varies over time, and she even considered variations of it that would be applicable to non-human animals if we spoke of ‘desire’ instead of ‘will’; see Bierson and Schwenkler (2022, 329n40). Williams (2006a, 98) also notes that the exact contours of the concept of the voluntary vary with the purposes to which it is put.

¹³ See Williams (1995n; 2005b, 271; 2006a).

conceptions that are activated on different occasions, depending on what kind of case one is considering.¹⁴

But the needs-based approach invites us to start further back, from a point of view that allows us to ask, more generally and with greater detachment from any particular context of application, why we would need *anything like* the concept voluntary. Why are we not better off leaving it among the ruins of the theory of volitions?

8.2 Making Sense and Knowing What to Expect

As its etymological roots in *voluntas*, the Latin word for *will*, indicate, the distinction we draw between the ‘voluntary’ and the ‘involuntary’ nominally separates actions that are attributable to the agent’s will from actions that are not attributable to the agent’s will, or attributable to it only to a lesser degree. But what characterizes the actions on each side of this distinction?

A helpful place to start is the observation that a user of the voluntary/involuntary distinction typically discriminates between different ways in which actions can relate to the intentional (i.e. cognitive and conative) states and deliberations of the agent. A voluntary action *expresses* those intentional states and, to the extent that the agent deliberated, the deliberation of the agent, because the agent has *shaped the action to fit them*. In an involuntary action, on the other hand, that expressive connection is typically weakened or absent altogether.

One basic and ubiquitous way in which an action can lose this expressive connection to the agent is by being unintentional. The things I do unintentionally tend not to reflect my intentional states, because they are actions I did not get a chance to shape to the intentions I formed in light of those intentional states. I might unintentionally delete a file I absolutely want to keep, for example. Indeed, I might unintentionally delete the file by intentionally updating my software. Every action is amenable to several correct descriptions, so that its being intentional under one description does not preclude its being unintentional under other descriptions. In Davidson’s canonical example, my action of flipping the switch can be correctly described as my turning on the light (a description under which it is intentional), but it can equally correctly be described as my

¹⁴ See Mele (2017, 137) for a such an occasion-sensitive account of the concept of free will.

alerting the prowler (a description under which it is unintentional).¹⁵ Hence, any action has multiple aspects, only some of which are intentional, and every action admits of descriptions under which it is unintentional. What the intentional/unintentional distinction discriminates between are therefore not actions *tout court*, but *aspects* of actions.

Yet there are various other ways in which the expressive connection between agent and action can be weakened.¹⁶ Consider reflex muscular contractions or automatisms, for example; or things done in a state of somnambulism or hypnosis; or in a state of intoxication, drug withdrawal, delusion, extreme incident passion, or insanity; or when muscular control is abruptly impaired by disease (e.g. a stroke, epilepsy, or Sydenham chorea); and what about things done under duress, or as a result of physical compulsion by someone else? These seem to express another person's intentional states and deliberation rather than the agent's. In each of these cases, there is thus some measure of *dissociation* of the action from the agent's intentional states and deliberation. But what concerns drive people to become conceptually sensitive to this dissociation?

The beginnings of an answer can be gleaned from Williams's argument that the concept of the voluntary is put within conceptual reach already by two distinctions that human beings everywhere can hardly avoid drawing: (i) the distinction between acting *intentionally* and acting *unintentionally*; and (ii) the distinction between acting *in a deliberatively normal state of mind* and acting *in a deliberatively abnormal state of mind*. According to Williams, there are very strong practical pressures on individuals living together to be sensitive to both of these distinctions, because their practical necessity follows already from some 'universal banalities' (1993, 55). By combining these two distinctions, he suggests, we already arrive at the concept of the voluntary as 'the idea of an intended aspect of something done in a state of mind that is deliberatively normal' (1995e, 242).

The distinction between things done intentionally and things done unintentionally is one that human beings everywhere have reason to use, according to Williams, because it is required to *make sense* of actions and to *know what to expect* from people—and these are things that human beings can hardly avoid being concerned to achieve. Williams

¹⁵ Davidson (2001, 4–5).

¹⁶ For a list of textbook examples from criminal law, see Hart (2008a, 95–6).

offers a vivid illustration from the *Odyssey*: as Odysseus and his son, Telemachus, confront Penelope's suitors, they are alarmed to find that the suitors are handing out weapons, even though Telemachus was supposed to have hidden away all their weapons in a storeroom. Odysseus angrily wonders who opened the storeroom, and Telemachus shamefacedly admits that the mistake is his—he did not look what he was doing and left the door of the storeroom ajar, and someone must have been a better observer than he was.¹⁷ Telemachus is clearly discriminating here between aspects of what he did that were *intentional* and aspects that were *unintentional*: it was he who left the door ajar, but he did not *mean* to—a subtle distinction in itself, but one which in this case marks the difference between fighting *with* Odysseus and fighting *against* him: had Telemachus intentionally left the door ajar, the implication for Odysseus would be alarming, suggesting that Telemachus was not in fact on his side.

This shows that even if Homer lacked a direct equivalent of our word ‘intention’, he had the *concept* of intention—not because we are disposed to draw on this concept in describing the situation, but because Homer and his characters themselves make distinctions which can only be understood in terms of that concept.¹⁸ Moreover, it would be surprising if they did not draw *some* distinction along these lines, because sensitivity to which aspects of an action are intentional is crucial to understanding what kind of action it is, and what to expect from one who intends such a thing in such a situation.

The concept of doing something intentionally, which enables one to discriminate between intentional and unintentional aspects of an action, is thus one that we are bound to have reason to use, because it is crucial to determining the significance of other people’s actions for us—which, as the case of Odysseus illustrates, can make the difference between life and death. Everywhere, human beings are concerned to make sense of action, if only to understand what exactly happened and what it means, and that requires understanding which aspects of an action were intentional and which were not. This is the inverse of Anscombe’s (1957, §21) point that the concept of intention would not exist if human beings took no interest in each other’s reasons for acting: we can take the fact that human beings everywhere can hardly avoid taking an interest in each other’s

¹⁷ See Williams (1993, 50). My discussion of Williams’s Homeric examples in this paragraph and the next draws on Queloz (2022b).

¹⁸ See Williams (1993, 50–51).

reasons for actions as being itself a reason to expect that they will turn out to possess the concept of intention.

The other distinction that human beings everywhere need to be conceptually sensitive to, according to Williams, is the distinction between things people do in what is for them a *deliberatively normal state of mind* and things they do when they are in what is for them a *deliberatively abnormal state of mind*. Any interpreter of human action has reason to look not only at which aspects of an action were intentional, but also at the action's relation to the agent's more settled concerns—concerns that, unlike momentary whims and impulses, relate to, and are possessed for, longer stretches of time. In particular, the interpreter must ask to what extent the agent was in a position to *shape the action to* those more settled concerns. If the agent was not in such a position, the action is defective in that it is dissociated from the agent, since the action fails to express and reflect the agent's more settled concerns. Actions performed in a state of somnambulism or under hypnosis are clearly defective in this way. But the defect may also lie in the fact that the agent's deliberation was severely skewed or entirely suppressed, as when the agent is in the grip of extreme passion.

Of course, there may not always be a clear line between deliberatively normal and abnormal states of mind, and the idea is not that we can specify, in universal and evaluatively neutral terms, which actions are defective or which states of mind are deliberatively abnormal. Different societies will elaborate the distinction between normal and abnormal states of mind differently, depending on what other concerns and conceptual material they bring to it. Some ways of drawing of the distinction may seem quite alien to us now. Here also Williams finds a vivid example in Homer: Agamemnon, who seized Briseis from Achilles, did so intentionally; but, as Agamemnon later explains, he was in an abnormal state of mind when he did so. He was not his usual self because the gods had cast what the Greeks called *atē* on his wits, temporarily engulfing him in a state of delusion and blind folly. By appealing to divine interference, Williams observes, Agamemnon is 'dissociating the action from himself' (1993, 54).

As differently as the distinction will be drawn in different societies, however, the key point is that human communities are bound to draw *some* distinction along these lines, because they are bound to take an interest in whether actions stand in regular relations to agents' more settled concerns. Like the distinction between intentional and

unintentional aspects of action, some such distinction between deliberatively normal and abnormal states of mind is also indispensable to making sense of what happened and to knowing what to expect from people, because the fact that an action was performed in an abnormal state of mind affects the meaning of the action, and renders the action less indicative of how the agent will behave in the future than actions that the agent had a chance to tailor to his or her more settled concerns. Things done in a deliberatively abnormal state of mind are not necessarily done *unintentionally*; but the state of mind from which the action flows interferes with the agents' capacity to shape their actions to their more settled concerns—either directly, by inhibiting their capacity to act on those concerns, or indirectly, by incapacitating the agents from deliberating properly and preventing these concerns from finding their usual expression in deliberation. Either way, the resulting actions are not representative of more settled concerns.

We thus have two concerns that give any human beings, insofar as they conceive of each other as performing actions, compelling reasons to apply something like the intentional/unintentional distinction and the normal/abnormal-state-of-mind distinction to those actions: the retrospective, hermeneutic concern to make sense of what happened, and the prospective, prudential concern to know what to expect from people going forward.

Once these two distinctions are in place, Williams argues, the concept of the voluntary is already within reach. For, with these two distinctions in place, we already have all the conceptual material required to construct the following notion of the voluntary:

A φ -s fully voluntarily if φ -ing is an intentional aspect of an action that A performs in a deliberatively normal state of mind.¹⁹

On this account, ‘voluntary’ action merits the term, i.e. is attributable to the agent’s *voluntas*, if and to the extent that it is intended by the agent in a deliberatively normal

¹⁹ This is my preferred reconstruction of Williams's notion of the voluntary. The gradability marked by the ‘fully’ comes out in one formula he uses: ‘an agent does X fully voluntarily if X-ing is an intentional aspect of an action he does, which has no inherent or deliberative defect’ (1995*l*, 25); see also Williams (1995*l*, 33n8; 2005*c*, 80n8; 2006*a*, 107). But my reconstruction leans more heavily on the formula he relies on most of the time: “A does X voluntarily” is equivalent to “A does X intentionally in a normal state of mind” (2006*e*, 120). See also Williams (1993, 66; 1995*g*, 73; 1995*n*, 495; 2006*a*, 107).

state of mind.

Three features of this conceptualization of the voluntary are worth highlighting. First, it is *gradable*: the closer an action is to a fully intentional action done in a fully normal state of mind, the more fully voluntary it is. Such a concept of the voluntary enables one to accept that every agent is inextricably enmeshed in a weave of contingent forces in which even the most paradigmatically voluntary action remains, in some respects, in the grip of influences beyond the agent's will. As Williams puts it: 'One's history as an agent is a web in which anything that is the product of the will is surrounded and held up and partly formed by things that are not' (1981c, 29). If a fully voluntary action is neither more nor less than an intentional action performed in a normal state of mind, then it is an action that may not have been under the control of the agent's will *in every respect*, but that was still *as much* under its control as actions ever are, and a great deal more so than an action done unintentionally or in an abnormal state of mind.

Secondly, and crucially for our purposes, this is a *superficial* concept, because it gives little attention to the detailed causal underpinnings of what it picks out. It licenses the ascription of voluntariness merely based on easily observable features of an action and its more immediate or proximal causes. Unless the notions of *intention* or of a *deliberatively normal state of mind* have themselves been tendentiously theorized to this end, such a concept does not commit its users to thinking that, at the deeper level of the ultimate causal underpinnings of actions, there must be a categorical difference between the metaphysical nature of what causes voluntary actions and the metaphysical nature of what causes involuntary bodily movements. Engaged users of this superficial concept of the voluntary can freely grant that whatever is a product of the will is inextricably tied up with what is not, and that the voluntary/involuntary distinction is not ultimately one whose real boundaries await discovery through a more detailed investigation of the causal processes underlying action.

Thirdly, although the applicability of this superficial concept of the voluntary remains insensitive to, and therefore tells us little about, the action's deeper aetiology beyond what most proximally caused it, this does not mean that a superficial concept of the voluntary *excludes* there being such ulterior explanations. It leaves room for the thought that *A* really φ-s voluntarily, but does so because *A* has been socialized a certain way, or is in the grip of an ideology. Indeed, as far as ideology critique is concerned, this compatibility of a

superficial notion of the voluntary with ulterior explanations is not a flaw, but a feature—for, as Étienne de La Boétie's 1576 'Discourse on Voluntary Servitude' (2016) already intimated, the real question for ideology critique is often not whether voluntariness is really servitude, but why the servitude really is voluntary. That is how Michael Rosen (2013) understands ideology critique, for example. If critical reflection on how *A* came to want to φ in the first place reveals some radical tension, it will not be a tension between φ 's claim to being voluntary and the fact that *A*'s wanting to φ admits of ulterior explanation, but rather a tension between that explanation and *A*'s understanding of what is really in *A*'s interest. (By contrast, *deepened* concepts of the voluntary really *are* incompatible with many ulterior explanations of how *A* came to intend to φ in the first place, which is why, as we shall see, deepened concepts of the voluntary are operative in sceptical challenges to free will).

Williams's view of the concept of the voluntary thus lies at the other extreme from the view that we can safely retire it. As he himself summarises his argument, the upshot is that the 'idea of the voluntary', far from being an idle residue of the theory of volitions, 'is inherent in the concept of action' (1995e, 247n4). He deems it 'reasonable to think that if we are to have the concept of an action, we must have the concept of a voluntary action' (1995e, 242).

This rather strong way of putting it might be taken to suggest that the 'must' in question is a matter of conceptual necessity—that the concept of the voluntary is necessarily already contained in the concept of action. But the nature of the argument that Williams actually offers suggests a weaker reading, for that argument appeals to *practical* necessity: to the conceptual need to develop certain distinctions, given the concept of action together with certain highly general concerns and circumstances: insofar as creatures like us conceptualize some of the things we bring about as actions, we are nearly bound, given our concerns to make sense of those actions and to know what to expect of people who so act, to develop the conceptual wherewithal to discriminate between different ways in which those actions relate to the intentional states and deliberations of the agents. And once these distinctions are in place, Williams contends, the concept of the voluntary is already within conceptual reach. This is how I think we should gloss Williams's otherwise perplexing claim that 'if we are to have the concept of an action, we must have the concept of a voluntary action'.

But, *pace* Williams, the mere fact that the concept of the voluntary is within the conceptual reach of a set of needful distinctions does not yet show that the concept of the voluntary is itself needed. It merely shows that we need a set of concepts from which the concept of the voluntary *can* be reached. But having a capacity in principle is one thing; being inclined to realize it in practice is quite another. For example, possession of the concepts *north* and *car* does not by itself guarantee that one will also be an engaged user of the concept *northcar*, which singles out cars pointing north for special treatment—though this possibility is intelligible to us as users of the concepts *north* and *car*, and might be said to be within the conceptual reach of a set of needful distinctions, we do not actually think and structure our affairs in those terms. The engaged/disengaged distinction again proves useful here: the set of concepts we can make disengaged use of is vastly larger than the set of concepts we have reason to make engaged use of. Showing that a concept falls into the former set is insufficient to show that it also falls into the latter.

So why did we actually yoke together what is done intentionally and in a normal state of mind under one concept and dignify it with its dedicated linguistic expression? Why did we form a new concept by drawing together just these properties and systematically differentiating in practice between actions that combine them all and actions that do not? Is it really a *new concept*, as opposed to a new word indicating the harnessing of two old concepts to new ends? To answer these questions, we must show that there is a conceptual need specifically *for* the concept of the voluntary. And to do that, we must understand what additional concerns fuel our interest in this particular grouping of properties and lend it its significance.

Even apart from these explanatory gaps in Williams's picture, however, there is a more fundamental reason why additional concerns need to be brought into the picture. For the concerns that Williams identifies as lying at the root of the concept of the voluntary, which I suggest are better described as lying at the root of its constituent notions, remain geared towards *predictability*: they are a matter of understanding people's actions with a view to knowing what to expect from them in the future. But if this combination of concerns were allowed to dictate the contours of the concept of the voluntary on its own, unchecked by additional concerns, the resulting voluntary/involuntary distinction would be hard to recognize. The notion of a deliberatively normal state of mind would be driven to become a quasi-statistical notion geared towards predictability alone, and

actions would count as voluntary, roughly, to the extent that they were likely to recur. This seems to yield the wrong result. It is nothing like the concept of the voluntary we know, which seems to be insensitive to mild untypicalities and not focused primarily on predictability.

Additional concerns must therefore be factored in if we are to identify reasons to preserve anything like the concept of the voluntary we have. The concepts acting as its building blocks may basically answer to the concerns to make sense of actions and know what to expect from people. But in the concept of the voluntary, this preoccupation with predictability is clearly balanced and checked by additional concerns, which pull the resultant concept of the voluntary away from a single-minded focus on what to expect.

8.3 Fairness and Freedom

One concern animating the use of the concept of the voluntary, I submit, is a concern for *fairness* in the attribution of responsibility, in the sense in which to carry responsibility for something is to be an appropriate subject of praise or blame with regard to it. Typically, one is treated as being blameworthy or guilty *only if* one acted voluntarily, and the observation that one acted involuntarily *excuses* one from being held responsible in this way, or at least acts as a mitigating circumstance. That arrangement tends to be mirrored in criminal law through what is sometimes called the ‘voluntary-act requirement’²⁰

Of course, certain types of cases call for compromises with more basic concerns, like the concern to avoid inherently dangerous situations (e.g. dangerous driving). In those cases, legal responsibility or liability may be ‘strict’, which is to say independent of the agent’s intentional states. Tort law also offers many examples of the voluntary-act

²⁰ See Moore (2010, 5) and Saunders (1988). The voluntary-act requirement is related to the *mens rea* doctrine (*actus non facit reum nisi mens sit rea*—the act does not make one guilty unless there is a guilty mind, which is to say a culpable intent; though there are also cases where negligence is enough for criminality). See Hart (1963, 40; 2008a, 90–92; 2008b, 36; 2008d, 14) for a discussion of the connection between the voluntary-act requirement and the *mens rea* doctrine; Duff (2004) argues that the voluntary-act requirement is a further requirement, which he articulates as *mens non facit reum nisi actus sit reus*—the mind does not make one guilty unless there is a guilty act, where ‘act’ is explicitly understood as ‘voluntary act’.

requirement being overridden by other considerations: when real damage has been done by something one involuntarily brought about and some response is called for—because *someone* needs to pay for repairs, for instance—one might end up being held responsible for something one brought about even if one strained every nerve to keep it from happening.

Generally, however, liability tends to be made conditional on voluntariness, and involuntariness is treated as being exculpatory. The law identifies exculpatory or mitigating circumstances under such headings as ‘accident’, ‘mistake’, ‘provocation’ into a passion causing the agent to lose self-control, or ‘insanity’.²¹ The characterization of voluntariness given above fits these excusing conditions. The former two correspond to the requirement that voluntary action be intentional, and the latter two to the requirement that voluntary action be done in a deliberatively normal state of mind. But, as H. L. A. Hart notes:

These psychological elements are not *in themselves* crucial although they are important as aspects of responsibility. What is crucial is that those whom we punish should have had, when they acted, the normal capacities, physical and mental, for doing what the law requires and abstaining from what it forbids, and a fair opportunity to exercise these capacities. (2008c, 152)

As Hart indicates, it is notably the concern for fairness that provides a rationale for rendering attributions of responsibility sensitive to the voluntary/involuntary distinction, because our capacity to do the right thing, given the kinds of circumstances we live in, is to a considerable degree hostage to contingent forces outside our control, so that holding people responsible for everything they did would be flagrantly unfair. It is, for example, an all too familiar fact that what one ends up having done, once the consequences of one’s intervention have unfolded, is largely a matter of luck, and even something done with the best intentions may issue in deplorable consequences. As a medieval proverb has it: once the flung stone leaves the hand, it belongs to the devil. A morality that allocated blame exclusively based on the eventual rather than the intended consequences of actions would turn blameworthiness into a plaything of contingent

²¹ See Hart (2008b, 31). As Hart also observes, much the same conditions are treated as invalidating civil transactions such as wills, contracts, or gifts.

forces, leaving it largely to sheer luck to decide whether one person attracted more blame than another. That would offend against our sense of fairness by putting the extent to which one attracted blame largely beyond the reach of individual control.

This conjunction of concerns, capacities, and circumstances calls for a concept that will help make attributions of responsibility fairer by focusing them on certain kinds of actions instead of leaving agents responsible for everything they do. In particular, there is a conceptual need for a concept that will contribute to responsibility being allocated on a fair basis. The concept of the voluntary fills this need. It serves to strike a balance between the social necessity of holding people to account for at least some of their antisocial behaviour and the concern to be fair in doing so by focusing responsibility on those aspects of action that are as much within their control as actions ever are, thereby mitigating unfairness in the attribution of responsibility. To base moral responsibility notably on the respects in which actions are voluntary is to focus moral appraisal precisely on those aspects of action that are as much as realistically possible within the control of the agent.²²

But there is also a second concern that animates the use of the concept of the voluntary: the concern for *freedom* as individual self-determination. In order to freely determine the course of their own lives, people need to be able to form determinate expectations as to how the moral code will affect them, and what kind of action they are likely to incur moral sanctions for. If one were subject to blame for everything one brought about, including what one brought about involuntarily, one's blameworthiness would become nearly impossible to anticipate. To leave agents any power to determine the course of their lives based on their sense of which actions will attract which kind of response, moral appraisal had better focus on those aspects of action over which agents have most control. The fact that attributions of responsibility are funnelled through the concept of the voluntary can thus also be understood as responding to a concern for freedom as individual self-determination.

This concern is particularly pressing when the code is *legal* rather than moral and the sanctions are state-enforced penal sanctions that constitute particularly severe threats to individual freedom. Individuals are concerned to determine the course of their own lives

²² This still leaves room for a moral authority that is exerted simply by what one has done—in cases that call for agent-regret as opposed to guilt, for instance; see Williams (1981c; 1993, 66; 2011, 196).

by acting with determinate expectations as to how the legal code will affect them and what they need to do and abstain from doing to stay on the right side of the law; but their capacity to anticipate all the eventual consequences of their actions is limited, given the kind of natural and social world they inhabit: how the eventual consequences of actions relate to the legal code is largely hostage to contingent forces—they cannot tell with certainty, for instance, whether their helping a person over the street today will be causally responsible for that person's death by food poisoning tomorrow. This array of concerns, capacities, and circumstances generates a conceptual need for a concept rendering penal sanctions more predictable by making liability conditional on those aspects of action that individuals can control and anticipate. The concept of the voluntary meets this need, and thereby serves the individuals' concern for freedom as self-determination. Without the concept—if people were liable simply for what they did or brought about—they would lose much of their power to determine and predict whether they stayed on the right side of the law. The law's emphasis on voluntariness thus expresses what Hart calls respect for the individual as a *choosing being* (2008b, 49).

Moreover, it is part of the point of many legal institutions—such as contracts, wills, gifts, and marriages—that they enhance individuals' ability to shape the future: they are legal tools enabling individuals to *lock in* certain outcomes.²³ For the law to interfere in individuals' lives in ways that they could not possibly foresee would therefore defeat one of the very purposes of the law. In the moral and even more so in the legal sphere, the idea of the voluntary thus performs a protective role, helping to respect and defend individual freedom against the claims of society.

8.4 Knowledge and Coercion

In a similar spirit, Hyman has recently argued that 'voluntariness is at root an ethical concept' that is 'designed for' the purpose of assessing a person's culpability, and that the concept is 'formed by negation', by 'excluding factors that exculpate' (2015, 76–77). Among the factors that are widely taken to exculpate are *ignorance* of what one is doing—sometimes called the 'knowledge condition' on responsibility—and doing something under *coercion* (e.g. at the point of a gun)—sometimes called the 'freedom condition' on

²³ See Hart (2008b, 29–30).

responsibility.²⁴ When Hyman proposes an *ex negativo* definition of voluntary action as what is ‘*not* done out of ignorance or compulsion’ (2015, 77), it is these two conditions he puts front and centre. This raises the question of how the understanding of voluntary agency as acting intentionally and in a deliberatively normal state of mind relates to these two conditions.

In the case of the knowledge condition, the relation is straightforward—the condition is already contained in the definition of the voluntary: the distinction between things done in full awareness of their nature and significance and things done out of ignorance is encapsulated in the requirement that one’s φ-ing must be an intentional aspect of what one does, for something can be an *intentional* aspect of what one does only if one is *aware* of it—if one *knows* what one is doing.²⁵ Of course, as recent debates over moral responsibility have made clear, the relevant notion of knowledge or awareness can be specified in various ways, depending on *how much* it requires one to be aware of and *what kind* of awareness it requires (does one need to *know* or merely *believe* it, and to do so *occasionally* or merely *dispositionally*?).²⁶ But just as human beings have reasons to discriminate between the intentional and the unintentional that are independent of their concern to allocate responsibility, they are bound to have an interest in *what an agent was aware of* when he or she acted, because that is already crucial to understanding the character of the action, the character of the agent, and the circumstances under which the agent would do something like that again. Did he know what he was doing when he intentionally φ-ed? That is, was he aware of the fact that he thereby ψ-ed? Or—what is sometimes rather more concerning—did he do it even though he did *not* know what it meant and what consequences it would have? Would he have done it anyway if he had known? If we tell him, will he refrain from doing it again? These are questions that human beings have an interest in asking already in virtue of being in the business of interpreting and shaping each other’s actions.

The freedom condition, by contrast, is not yet contained in that definition of the voluntary. Indeed, Williams, like Anscombe, is comfortable describing actions done

²⁴ See, e.g., Fischer and Ravizza (1998, 12–13) and Rudy-Hiller (2018).

²⁵ This is also how Williams (1995I, 23–26) understand the intentionality requirement.

²⁶ See Rudy-Hiller (2018) for a helpful overview.

under coercion as voluntary.²⁷ He emphasizes that decisions reached under coercion really are *decisions*, characteristically coming out of a process of deliberation issuing in the conclusion that one *has to* do something, because the coercive threat overrides other deliberative priorities.²⁸ This is not to deny that coercive circumstances can be exculpatory. It is merely to deny that every exculpatory circumstance must register as such by going into forming the voluntary/involuntary distinction. As Williams insists, ‘the topic of coercion is not part of the theory of action, but of the theory of freedom’ (1995*l*, 33n3).

But if, as I have suggested, a central need that the concept of the voluntary answers to is the need to separate, in the name of fairness and freedom, actions that are *more* from actions that are *less* under the agent’s control, then perhaps we do, after all, have reason to refine the notion of the voluntary we started out from to exclude coerced action. There is an undeniably important sense in which the agent who acts under threat of serious harm is deprived of control over the action, and the intentions expressed by the action are not really *the agent’s* at all, but the coercer’s. Like acting in an abnormal state of mind, this dissociates the action from the agent. And indeed, legal codes tend to treat the fact that an action was done under *duress* (coercion by the threat of serious harm) as an excusing condition; and Hart, for example, characterizes coerced actions as *involuntary*.²⁹

To better serve the concerns that fuel our interest in it, therefore, the notion of the voluntary should be understood more restrictively:

A φ -s *voluntarily* if and to the extent that φ -ing is an intentional aspect of an action that A performs uncoerced and in a deliberatively normal state of mind.

One might protest on Williams’s behalf that this introduces a moral element from the theory of freedom into what was supposed to be, in the first instance, a psychological concept in the theory of action, on a par with concepts of choice, decision, belief, or

²⁷ See Anscombe (2008b, 127) and Williams (1995*l*, 33n8). In an archival note, however, Anscombe notes that ‘voluntary behaviour is behaviour in respect of which the behaver is free. To be free is to be in a situation of possibility of determining something to accord with one’s will’ (Archive, Box 9, File 304, p. 1; cited in Bierson and Schwenkler (2022, 329n48)).

²⁸ See Williams (1995*c*, 5; 1995*l*, 33n3, 33n8).

²⁹ See Hart (2008*c*, 143–44; 2008*d*, 14).

desire.

But even the earlier, more permissive definition of the voluntary could hardly escape drawing on a morally laden understanding of what kinds of states of mind appropriately attract appraisals of responsibility. Moreover, if the reasons we have to use the concept of the voluntary in the first place, over and beyond the concepts of acting intentionally and in a normal state of mind, are *moral* reasons grounded in moral concerns, then it is only right that these moral concerns should also substantively *shape* the resulting concept of the voluntary. In discussions of thick concepts, the point is often made that a thick concept's extension is a function of our evaluative interests in deploying the concept, and not specifiable independently of them. One might argue that the same is true of the concept of the voluntary. When taken in isolation from the concept of the voluntary and the moral concerns that give us reasons to use it, its constituent notions, such as that of a deliberatively normal state of mind, are unlikely to have the same extensions as when deployed alongside other constituent notions in the service of moral concerns. This is why the concerns for fairness and freedom call for a genuinely new concept that is more than the sum of its constituent concepts. The notions that coalesce in the concept of the voluntary need to be understood in a certain way and be appropriately linked. Neither these constituent notions nor their appropriate link are specifiable independently of the moral concerns that rationalize and inform the concept of the voluntary.

At the same time, Williams and Anscombe are right to emphasize that the idea of the voluntary is, in the first instance, a concept in the theory of action.³⁰ It really is a psychological concept—only one that acts as a lynchpin between psychological and moral concepts, connecting concepts of action, deliberation, intention, and decision with concepts of praise, blame, responsibility, liability, and exculpation. What the concept of the voluntary picks out is a collection of properties of actions, which properties they have in virtue of their relation to the agent's deliberation and state of mind; but the *second-order reasons* for picking out just these properties and grouping them as constituents of a significant, reason-giving property—voluntariness—are *moral* reasons, and the inferential consequences of displaying or lacking that property are *moral* consequences. Hence, the idea of the voluntary marks a place where ideas about mind and action make contact with moral demands. In a hybridizing classification that will prove helpful in what

³⁰ Pace Hyman (2015, 76–77).

follows, we might say that the concept of the voluntary is a psychological concept that serves and is shaped by moral concerns.

Finally, the way in which our concerns for fairness and freedom each shape and call for a concept of the voluntary along these lines can also be drawn upon to vindicate the concept's authority along with our confidence in it. In showing that the concept does important work for us by meeting our needs and serving our concerns, one highlights reasons for us to continue to use the concept and heed the reasons it advertises to, thereby vindicating our confidence in the concept by showing it to be well-placed and reasonable. We have a conceptual need for the concept of the voluntary because we are concerned, in the name of fairness and freedom, to exempt individuals from some of their moral and legal responsibility for some of their involuntary acts. Of course, users of the concept do not necessarily consciously *aim* to serve a concern for fairness or freedom. Nor do these concerns figure in the *content* of the concept. But the needs-based approach allows us to show that we need the concept for these reasons all the same.

8.5 When Concerns Distort Concepts

What makes the case of the concept of the voluntary particularly interesting, however, is that it offers a vivid illustration not only of how concerns can shape concepts, but also of how they can distort them.

The upshot of the argument thus far has been that we need the concept of the voluntary in *some* form because we feel the need, for the sake of fairness and freedom, to exempt individuals from some of their moral and legal responsibility for some of their involuntary acts.

But the concern for fairness aims at making attributions of responsibility not just *fairer* than they would be without the concept of the voluntary, but *ultimately* fair. The concept of the voluntary we sketched goes some way towards meeting the demand for fairness, but by no means all the way. It still leaves a great deal of residual unfairness, because *whether one is in a position* voluntarily to do the right thing itself remains a matter of luck—a matter of one's inherited disposition, upbringing, and socialization, but also a matter of whether and how often one finds oneself in circumstances that make it especially hard to do the right thing. Consequently, one's capacity to voluntarily do the right thing remains partly a matter of constitutive and circumstantial luck.

Appraisals of responsibility funnelled through this concept of voluntariness are *fairer* than they would be if they completely ignored the voluntary/involuntary distinction, but they are not *ultimately* fair. They still fall foul of the demand for ultimate fairness that Michael Zimmerman forcefully expresses when he insists that ‘the degree to which we are morally responsible cannot be affected by what is not in our control. Put more pithily: luck is irrelevant to moral responsibility’ (2002, 559).

The moral concern for responsibility to be attributed on a fair basis thus exerts pressure on the concept of the voluntary to accommodate a demand for *ultimate* fairness. There is a pressure on the concept to fully isolate that which is a product and expression of the agent’s will from that which is a product of contingent—and hence not necessarily fairly distributed—forces. Much as the concept of the voluntary would be transformed into a quasi-statistical concept if its contours were allowed to be determined solely by a concern to predict behaviour, the concern for fairness, in and of itself, pulls the concept towards a form in which it ensures complete fairness.

The result is a kind of *moralization of psychology*: a warping of psychological concepts to serve moral concerns. Such a moralization of psychology need not always be problematic—there is nothing inherently wrong with tailoring a psychological concept to the moral concerns it ties in with. But the moralization of psychology becomes *problematic* when it goes further than that, and the psychological concepts are warped under moral pressure to the point where they become *psychologically unrealistic*, which is to say when they become inconsistent with the rest of what we take ourselves to know about human psychology and how the world works. Our moral concepts should be answerable to human psychology, not the other way round. Owen Flanagan enshrines this idea in his ‘Principle of Minimal Psychological Realism’: that ‘when constructing a moral theory or projecting a moral ideal’, one should make sure ‘that the character, decision processing, and behavior prescribed are possible, or perceived to be possible, for creatures like us’ (1991, 32); or, as Samuel Scheffler succinctly puts it, only a psychologically realistic morality can be a *human* morality.³¹ The risk, then, is that if the concern for fairness is allowed to dictate the contours of the concept of the voluntary without counter-pressure from other concerns, it will *distort* it into a concept that does not strike a helpful balance between our concerns.

³¹ See Scheffler (1992, 7–9).

This is again a point at which we can usefully turn to Williams, because the idea that psychology can become problematically moralized under the pressure of moral demands is a thread that runs through much of his oeuvre. Already in 1963, Williams makes the point—which he credits to Iris Murdoch—that it is an ‘evaluatively motivated picture of the mind’ which ‘sharply distinguishes between “reason” and “will”’ (1963, 136). In *Shame and Necessity*, he then suggests that it was Plato who first ‘ethicized psychology’ with his tripartite model of the soul, because he defined ‘the functions of the mind, especially with regard to action … at the most basic level in terms of categories that get their significance from ethics’ (1993, 160). In particular, Plato’s stark division between ‘rational concerns that aim at the good, and mere desire’ (1993, 42) enabled him to introduce a ‘featureless moral self’ (1993, 160) into his psychology, a locus of agency that remained uncontaminated by contingent desires.

What makes this an example of a problematically moralized psychology is not that it plays a morally significant role, or even that it draws on values—a realistic psychology, as Williams notes, need not be ‘value-free’; but it ‘leaves it open, or even problematical, in what way moral reasons and ethical values fit with other motives and desires, how far they express those other motives, and how far they are in conflict with them’ (1995i, 202).³² What a problematically moralized psychology lacks is precisely this *openness*: it closes off the very possibility of ineliminable conflict between the psychological and the moral. Psychological concepts are distorted to ensure that moral concerns can be met.

In the case of the psychological concept of the voluntary, this problematic moralization of psychology takes the following form: to enable attributions of responsibility to be made on an *ultimately fair* basis, the concept of the voluntary needs to make responsibility a function of something that meets both of the following conditions: (a) it must be something that is perfectly evenly distributed, so that every agent has an *equal opportunity* to do the right thing; and (b) it must be something that each agent has *total control* over, so that no contingent circumstances constrain or predetermine the agent’s capacity to do the right thing.

³² See also Williams (2006f, 78).

8.6 Deepening the Concept of the Voluntary

To meet this conceptual need for a way of thinking that makes responsibility a function of something that is perfectly evenly distributed and entirely within the agent's control, the concept of the voluntary must become the concept of the *purely* voluntary—of something that is pure of any contamination by the potentially unfair influence of empirical contingencies.

To achieve this, the concept of the voluntary must be *deepened* in a particular way, namely so that the locus of responsibility retreats deep enough into the agent to be sheltered from contingency. It is not enough for responsibility to track what one *wills* rather than what one contingently ends up having done, since one's capacity to align one's will with morality or the law is itself subject to constitutive and circumstantial luck—a matter of what dispositions one inherited and what challenges life presents one with. If the basis on which we allocate responsibility is to be ultimately fair, various contingencies *within* the agent need to be eliminated as well, to immunize responsibility against the contingent influence of natural endowments, socialization, education, and other biographical and historical circumstances.³³

Accordingly, the concept of the voluntary must be deepened in such a way as to cast voluntary action as *fundamentally distinct in nature* from involuntary bodily movements and *pure of contingency*, i.e. fully isolated from contingent forces, thereby grounding not just a distinction of degree, but a sharp, categorical distinction between the agent's will and the contingent forces that are external to it. If the agent's will were always, however slightly, under the influence of forces beyond itself, this would again introduce an element of contingency or luck—hence of potential unfairness—into all the agent's actions.

The theory of volitions offers a paradigmatic example of an attempt to deepen the concept of the voluntary in a way that promises to meet this set of demands. For what marks out someone like Descartes as operating with a deepened idea of the voluntary is not the mere thought that what turns a bodily movement into a voluntary action is its

³³ This point is at the heart of Williams's critique of the 'morality system' as I reconstruct it; see Queloz (2022b) and Queloz and van Ackeren (manuscript). See also Moore (2003, 2006a), Louden (2007, 110–11), and Russell (2013, 2018, 2019, forthcoming).

tracing back to an act of will or a volition,³⁴ it is the further fact that the act of will or volition is itself understood as something that is distinctively *pure of contingency*—something that lies beyond the reach of empirical determination by powers external to the acting subject. For Descartes, a volition is an action of the mind or soul; and while the soul can affect matter by affecting the pineal gland (which affects the animal spirits which in turn affect the muscles), the soul is not itself affected by anything beyond itself.³⁵

On such a theory, voluntary actions have their ultimate source in something that lies deep enough to be completely isolated from the blind play of chance. There is an echo here of the Stoics' attempts to shelter themselves from upsetting strokes of fate, and Descartes' theory of the will went hand in hand with an elaborate Neo-Stoic ethic that also treated the emotions as being fully under the will's control (he once urged Elisabeth of Bohemia to look at the bright side of her uncle's decapitation).³⁶ If voluntary action is rooted in something fundamentally distinct from the muck of contingent forces, this categorically distinguishes it from other happenings: it is not just *less* mired in contingency, but, at base, completely pure of it.³⁷

A more recent example of a deepened concept of the voluntary can be found in Benjamin Libet's experiments, in which subjects hooked up to an electroencephalograph were put in an optimal position to deliberate at their leisure about whether and when to flick their wrists. What motivated the experiments was precisely the perception—which indicates a deepened concept of the voluntary—that the mere fact that subjects carefully deliberated and acted only when they really meant to did not yet settle the question whether they *voluntarily* flicked their wrists. *That* was to be determined by having each

³⁴ This much could be granted by more recent action theorists like O'Shaughnessy (1973, 2008a, b) and Hornsby (1980), even though they develop accounts of acting and willing in terms of the notion of *trying to φ* that do not deepen the concept of the voluntary and are not threatened by the possibility that determinism might be true and do not require us 'to look back beyond the trying' (Hornsby 1980, 59).

³⁵ See Descartes (1996, III 372, XI 342). See also Kenny (1972), Alanen (2002), and Jayasekera (2016). As Williams notes, Descartes's attempt to explain how I can move my body at will in terms of 'a kind of internalized psychokinesis' whereby the pineal gland is the only part of the body that is directly responsive to the will has the uninviting consequence that 'the only part of my body directly responsive to my will is one which I cannot move at will' (2005b, 277).

³⁶ See Schneck (2019, 757n12).

³⁷ See Williams (2005b, 271).

subject note the time at which they formed the intention to flick their wrist and comparing it against the neurophysiological processes underpinning their conscious deliberation and action. If one conceives of voluntary action as having its ultimate or terminal causal source in a will that is distinct from and exerts control over neurophysiological processes,³⁸ this deepened concept of the voluntary licenses the inference from the observation that neurophysiological activity *precedes* conscious willing to the conclusion that the action in question is not voluntary.

This example also illustrates that by deepening the concept of the voluntary, one typically renders it more costly to use, requiring concept-users to dig deep into the aetiology of an action in order to determine whether or not the concept applies, because the *real mark* of voluntariness is to be found further back; and it illustrates that a deepened concept of the voluntary can manifest itself primarily *negatively*, i.e. through what is treated as foreclosing the application of the concept. People can be very clear about what does *not* count as the right kind of neurophysiological underpinning of voluntary action without being clear about what *would* count as the right kind.

Not all ways of deepening the concept of the voluntary must involve postulating influences coming from beyond the material or natural order. One might, for instance, maintain that an action counts as voluntary if and only if it *reflects nothing but character traits and dispositions that the agent voluntarily acquired*. Not only the action, but even the agent's *becoming* the kind of agent who is disposed to act in this way in that situation would then fully be the product and expression of the agent's will, undistorted by external influences. This does, however, presuppose that what character the agent came to develop was fully under the control of the agent's will.

One might also try to deepen the concept of the voluntary without this presupposition, by maintaining that an action counts as voluntary if and only if the agent acts *exclusively out of motives that are fully responsive to the agent's deliberation about what motives to have*—a type of account which admits that we do not, originally, acquire our

³⁸ Another interpretation—which is closer to that favoured by Libet himself—is to conceive of the will as the kind of thing capable of *blocking* or *vetoing* the neurophysiological processes resulting in action. That way of securing the agent's full control over actions also traces back to the Stoics, who thought of the *hegemonikon*, the rational and leading faculty of the soul, as a gatekeeper whose assent was necessary for impressions to issue in action.

motives voluntarily, but seeks to reestablish the full autonomy of the will through the idea that an agent's first-order volitions—e.g., *A*'s wanting to φ —can fully come under the deliberative control of her second-order volitions—e.g., *A*'s wanting the desire to φ to be her will. Here the idea of the voluntary is deepened through a picture of human psychology on which one's rational deliberation can exert total control over one's motives. Though it confines itself to the natural and material order, this also goes a long way towards deepening the concept of the voluntary in the way required, since it offers an account of the nature of voluntary action that categorically separates it from the influence of contingency. It does, however, presuppose that one can achieve total deliberative control over one's motives.

What all these different ways of deepening the concept of the voluntary have in common is that by locating voluntary agency beyond contingent empirical determination, they give the psychology of agency the right shape to hold out the promise of allocating responsibility and blame on an ultimately fair basis. The deepened concept of the voluntary promises to be a shelter from luck.

The supposed capacity of a deepened concept of the voluntary to remedy the world's unfairness might be thought to give us good reason to prefer a deepened concept of the voluntary over a superficial one. But, far from performing the same tasks better, a deepened concept of the voluntary ceases to do the work it did in its more superficial form. Though it promises to meet the concern for fairness better than the more superficial concept, that promise ultimately turns out to be illusory: it ends up serving our various concerns markedly less well and landing us in broader trouble.

8.7 The First Problem of Free Will

The source of that trouble is that if some psychological concept is problematically moralized under pressure from moral demands, this generates a pressure on the rest of our psychological concepts to support that moralized concept, and a corresponding pressure on our physical or metaphysical concepts to support such a psychology. The relevant dynamics can be captured in terms of a simple three-tier model. Think of our moral, psychological, and physical concepts (in the broad sense of 'physical' that includes biological, physiological, and metaphysical concepts, as it did in the ancient notion of *physis*) as organized into three vertically layered tiers, with the physical tier at the bottom,

the psychological tier in the middle, and the moral tier at the top. Each tier comprises concepts such as the following:

Moral concepts: moral responsibility, blame, praise, justice, fairness, freedom, etc.

Psychological concepts: voluntariness, will, trying, intention, choice, decision, action, belief, desire, etc.

Physical concepts: bodily movements, cause, effect, determinism, quantum randomness, etc.

This three-tier model takes us beyond the traditional two-tiered framing of the free will problem as a matter of whether our *moral* concepts are compatible with *physical* concepts such as determinism or randomness at the quantum level. When set against the three-tier model, the two-tiered framing of the problem in terms of the possibility of reconciling moral responsibility with determinism looks, as Williams puts it, like a ‘structural misconception’ (1995c, 6), for what the three-tier model brings out is that, *in the first instance*, ‘our ideas of blame and responsibility are answerable to an adequate psychology (rather than to generic worries about determinism)’ (1995d, 45n10). There are not one, but *two* points at which the question of the relation of one set of concepts to another set arises.

As a result, we face not one, but two problems of free will:

- (1) *The First Problem of Free Will*: How can our physical concepts be reconciled with our psychological concepts?
- (2) *The Second Problem of Free Will*: How can our psychological concepts be reconciled with our moral concepts?

This in turn helps us to see that securing a fit between moral and psychological concepts by deepening the concept of the voluntary comes at a price: it exerts a corresponding pressure on psychological concepts that is likely to bring them into tension with physical concepts. This is what generates the first problem of free will: if the will expressed in voluntary action is to be a force that is entirely pure of any conditioning influence by unfairly distributed contingencies, how are we to make sense of such a force in physical terms?

To resolve the tensions that problematically moralized psychological concepts generate in relation to physical concepts, the moralization must *seep through* or *extend to* the physical. That is, one's conception of nature or its metaphysical underpinnings must be adapted to meet the moral demand for some deepened concept of voluntary agency that can provide an ultimately fair basis for responsibility. Aristotle's biology, with its idea that the proper natural development of the human animal issues in virtue, can be thought of as exemplifying a conception of nature that is tailored to ensure that moral demands can be met. Similarly, Plato's dualistic metaphysics of the soul and his theory of the Forms support his tripartite psychology in just the way required to secure the realizability of his moral ideals.

But perhaps the best example of a moralized metaphysics tailored to support moral demands that are specially focused on the concept of the voluntary is Kant's postulation of the noumenal realm as the locus of the transcendental subject's unconditioned will.³⁹ On the Kantian picture, moral goodness pertains to the good will rather than to its consequences, and the good will is understood in such a way that one's capacity to exhibit it remains unconditioned by contingent empirical circumstances: only the will to do one's moral duty counts as good in itself, and only if it is motivated by the rational insight that it is one's duty rather than by the contingent natural inclinations that some people have and others lack. Since the rational faculty is something which, on Kant's view, is perfectly evenly distributed, and since the noumenal realm provides a place for a source of voluntariness that lies beyond empirical determination, this picture offers solace against unfairness and a true shelter from luck. The only thing that ultimately matters—the goodness of the unconditioned will of the noumenal self—is completely within one's control, for it depends only on whether one chooses, from motives everyone equally has anyway, to align one's unconditioned will with one's moral duty. Accordingly, Kant reassuringly proclaims: 'To satisfy the categorical command of morality is within everyone's power at all times' (2015, 33).

On a picture on which a moralized psychology finds support in physical concepts, there can be a harmonious fit between the fundamental nature of reality, the operations of human psychology, and the demands of morality. There is considerable attraction in

³⁹ Though some have insisted that even Kant ultimately does not deny that moral action is subject to contingent empirical determination in certain respects; see Heyd (1997) and Hartman (2019).

such a picture. It holds out the promise that everyone will be offered an equal opportunity to live up to the demands of morality: if the moral shape of one's life depends exclusively on what one voluntarily does; if the voluntary/involuntary distinction is applied at the deeper level of the purely rational self rather than at the superficial level of the contingently constituted empirical self; and if one's metaphysical view of things supports the existence of such a deeper level; then the moral shape of one's life really can be entirely within one's control and reassuringly sheltered from luck.

But, on a thoroughly disenchanted conception of the world, it is far from clear that a psychology moralized along these lines can find the support it requires in our physical concepts. The story of the rise of modern science is also the story of how the moralization of our physical concepts was reversed: physical concepts were *unmoralized*, in the sense of rendering these concepts more emancipated from moral demands and less accommodating of them. As a result, the deepened concepts of the voluntary at work in the theory of volitions or in the Kantian theory of the unconditioned will no longer neatly dovetail with our physical and metaphysical concepts, and it comes to look as though the concept of a voluntary act is never fully instantiated. Our modern physical ideas leave no room for pure volition or an unconditioned will. Hence, the unmoralization of physical concepts creates a tension with moralized psychological concepts and renders them problematic, by making it hard to see how human agency can live up to that conception of it.

The effect of combining in people's minds a deepened concept of the voluntary with physical ideas that deny its instantiation is not *necessarily* to produce the conviction that no act is ever voluntary. People are quite capable of holding on, at once and without realizing it, to two sets of ideas that are in tension with each other. But it does *set them up* to conclude, when they reflect on the voluntariness of a particular action and inquire more deeply into its aetiology, that the action was not *really* voluntary, because it reflected the agent's genetic predispositions, or upbringing, or some other circumstantial factor. On this basis, they may come to realize that something similar holds true of every action they can point to. *Pace* the compatibilists, they would then not be mistaken about the incompatibility of their physical ideas with their idea of voluntariness, but would correctly have grasped the implications that a deepened concept of the voluntary has when set against a naturalized or disenchanted conception of the world.

If one's concept of a voluntary action requires the action to be pure of contingency or luck in every respect, sufficiently close scrutiny will thus eventually disqualify any action when set against a disenchanted conception of the world, leaving the concept's extension empty and denying it any applicability to our actual experience. And insofar as the concept loses its applicability to our actual experience, it also fails to mark out any actions as appropriate objects of praise or blame. Perhaps some concepts are worth having even if they are never satisfied, but insofar as the concept of the voluntary earns its keep by rendering us sensitive to *differences* between human actions of the kind actually to be met with, it is certainly not one of those concepts. Far from delivering on the promise of ultimate fairness, its failure to carve off any actual action as voluntary incapacitates the concept from serving our concerns for fairness and freedom at all.

It follows that if the concept of the voluntary (along with the appraisals of responsibility that build on it) is to do any work for us in a world in which every action displays an element of contingency, its extension within that world had better not shrink to the point of becoming empty: the contrast between the voluntary and the involuntary had better remain an internal contrast *within* our experience. If our moral and legal concepts are to be applicable to us contingently determined agents, the concept of the voluntary that guides their application must accommodate some measure of contingency. It needs to retain the ability to draw a contrast within the world we live in, by contrasting, not conditioned with unconditioned actions, but different kinds of conditioned actions—that is to say, it needs to draw the voluntary/involuntary distinction *within* a range of actions that are *all* adulterated with contingency to some degree, but adverting instead, and rather more helpfully, to the differences in *how much* contingency is involved them. It thus renders us, and our practices of responsibility attribution, sensitive to the *degree* to which a conditioned action is attributable to the agent's will.

This is something that a concept deepened to deliver ultimate fairness could not do for us, because it could not be gradable. To meet the demands of ultimate fairness, it would have to be a dichotomous rather than continuous concept, categorically separating voluntary action from every other kind of action or happening at the deepest causal level. And because it could not be gradable, it would level all differences between variously conditioned actions. By its light, any action that was not completely unconditioned and

purely voluntary would appear as externally determined and dissociated from the agent as the next.

If we refrain from deepening the concept of the voluntary, by contrast, it *can* be a more nuanced, gradable notion: the less an action is the product of forces external to the agent's will, the more voluntary it is, and the more the agent deserves to be held accountable for it. Such a concept of the voluntary enables us to accept that even the most fully voluntary action reflects influences beyond the control of the agent's will while still retaining its differential applicability to our actions. It thereby serves the need to hold people accountable for at least some of their actions while also serving the concerns for freedom and fairness as far as realistically possible.

The kind of concept of the voluntary we need, then, is one that balances the moral concerns for freedom and fairness against the practical concern to hold people accountable while being realistic about the pervasiveness of contingency: it should render us sensitive to differences in the degree to which actions are under the control of the agent's will while retaining the wide applicability of our regulatory practices of responsibility attribution by accommodating the fact that even the most voluntary actions are still conditioned by contingent factors lying beyond the agent's control. As long as it remains superficial, the concept of the voluntary does important work for us by enabling us to realize our concerns for fairness and freedom in our appraisals of responsibility as far as the pervasiveness of contingency will allow.

It might be objected that as long as it remains superficial in this way, the concept of the voluntary does little to answer traditional worries about free will. But the point is that it also avoids raising those worries in the first place, and thereby avoids threatening the legitimacy of our appraisals of responsibility across the board. It is only insofar as the concept of the voluntary is deepened that the unmoralization of physical concepts generates what we identified above as the *first* problem of free will, for only then will the concept of the voluntary come into conflict with physical ideas such as determinism. In its superficial form, the concept generates no such conflict, as suggested for instance by Brian O'Shaughnessy's reconciliation of causal determinism with voluntary action.⁴⁰ That is why Williams, who explicitly endorses O'Shaughnessy's reconciliation,⁴¹ writes in

⁴⁰ See O'Shaughnessy (2008a; b, esp. chs. 11 and 17).

⁴¹ See Williams (1995a, 579; 1995c, 8).

his study of the pre-Socratic Greeks that ‘there is a problem of free will only for those who think that the notion of the voluntary can be metaphysically deepened’ (1993, 68). Williams’s turn to the pre-Socratic Greeks is driven by a desire to recover, or remind us of the value of, a superficial concept of the voluntary as yet unmarked by the moral pressure to deepen it.

8.8 The Second Problem of Free Will

Where does this leave us? In the end, it still leaves us facing the *second* problem of free will: the problem of how to reconcile our psychological with our moral concepts. For if the compatibilists are right about the first problem of free will, there is also an important sense in which the incompatibilists are right about the second problem of free will. Even if we thoroughly unmoralize our psychological concepts and succeed in reconciling them with our physical concepts, there remains a tension between our psychological concepts of human action and our moral concepts.

At least, there remains a tension between the superficial concept of the voluntary and the host of moral concepts we have that encode the demand for morality to be pure of contingency: the concept of a *moral reason* or a *moral motive* as something that contrasts sharply with other kinds reasons or motives in claiming to be completely independent from contingent personal attachments, inclinations, loyalties, and projects; the concept of a *moral action* as something that only counts as such if it is motivated by reasons that are moral in this pure sense, and not by dispositions or motives one contingently has; the concept of *moral obligation* as something one is under anyway, whatever one’s contingent motivations and commitments; and the concept of *moral blameworthiness* as something that only attaches to the voluntary breaking of moral obligations. As long as the concept of the voluntary that underpins these concepts is itself pure of contingency, morality presents itself as something that transcends luck. Life may be unfair in other respects, but determining the moral shape of one’s life, these concepts encourage one to think, is something everyone gets a fair shot at.

The difficulty, however, is that as long as these moral concepts shape our thinking, we cannot simply accept that the concept of pure voluntariness cannot be satisfied and opt for a superficial concept of voluntariness instead. For if we do this against the background of these luck-intolerant moral concepts generating certain normative expectations about

what shape the world can properly have if free agency and its concomitant moral import are to have a place in it, the eventual result must be a disillusioned scepticism about the very possibility of moral agency: a view on which no act is every truly voluntary, and the noble edifice of morality turns out to have been built around an illusion.⁴²

There is thus a danger in admitting to what extent contingency pervades human action while retaining the concepts embodying the expectation that morality should be free of luck, since those concepts blind us to the forms of freedom and moral motivation that really are to be found in the world we live in. These concepts leave us, as the only contrast to their purified conceptions of moral agency and blame, a flattened picture on which everything is contingent inclination, compulsion by brute causal forces, and manipulation of those same forces. Consequently, no action is ever completely pure of contingent influences in the way morality requires, and morality never really gets a foothold in our world.

This is, of course, one of the central complaints that both Nietzsche and Williams levelled against the Stoic, Christian, and Kantian elaborations of morality.⁴³ They see the presupposition of pure voluntariness as presenting an insuperable hurdle for compatibilism. That is why Williams, despite his seemingly complacent dismissal of the free will problem in the passage quoted at the beginning of this chapter, is not a compatibilist.⁴⁴ He does not think that, as things stand with our moral concepts, the second problem of free will admits of a compatibilist solution:

⁴² On sceptical challenges to free will, see Pereboom (2001), Waller (2011), and Levy (2011). As Russell (2017b, xiv) points out, such sceptical challenges are now taken more seriously than a few decades ago, when the main debate was still between libertarians and compatibilists.

⁴³ On Nietzsche's critique, see Owen (2007), Leiter (2015), Clark (2015), Queloz and Cueni (2019), Richardson (2020), and Reginster (2021). On Williams's critique, see Moore (2003, 2006a), Louden (2007, 110–11), Russell (2013, 2018, 2019, forthcoming), Queloz (2022b), and Krishnan and Queloz (2022); on its extension to Stoicism, see Queloz and van Ackeren (manuscript).

⁴⁴ *Pace* Leiter's (2022, 30) characterization of Williams as a compatibilist, both Williams and Nietzsche are thus incompatibilists about deepened concepts of the voluntary, and derive much of their critical leverage over morality from that fact. While Williams *is* a compatibilist about a superficial concept of the voluntary, however, Leiter (2019a, b) reads Nietzsche's philosophical psychology as leaving intentions and reasons for action so epiphenomenal in comparison to the unconscious drives that Nietzsche casts as the real determinants of action that this would make Nietzsche an incompatibilist even about superficial concepts of the voluntary.

Can the reconciling project succeed? Between determinism (or as much naturalistic explanation as you like), and relevant psychological concepts, yes. Between both of these, and the ethical conceptual scheme, no, not as it stands. (Williams 1995c, 19)

As Williams sees it, the constellation of moral concepts we inherited ‘makes people think that … without its utter voluntariness, there is only force; without its ultimately pure justice, there is no justice’ (2011, 218)—when ‘in truth, almost all worthwhile human life lies between the extremes that morality puts before us’ (2011, 216).

The root cause of this discrepancy, I have suggested, is a distortion of the concept of the voluntary by an overwhelming concern for fairness—a concern for ultimate fairness that puts pressure on morality to offer a shelter from luck. What makes it a *distortion* rather than a helpful tailoring of the concept to the need engendered by a noble human concern is the concept’s relation to the balance of our concerns. The concern for fairness is not our sole concern, and it must be balanced against other pressing concerns, including the concern for freedom and the concern to regulate behaviour by holding people accountable for some of their actions. But even on the terms of the concern for fairness alone, a deepened concept of voluntariness fails to serve that concern, because far from delivering ultimate fairness, it ends up failing to deliver even the more modest gains in fairness that are really to be had by means of this concept alone. In a world pervaded by contingency, a deepened concept of the voluntary, correctly applied, can still *express* the concern for fairness, but it cannot serve it.

Once one accepts that the demand for pure voluntariness cannot be met, there are two ways one can go.⁴⁵ One can renounce the deepened concept in favour of a more superficial and luck-tolerant one, and allocate responsibility on that basis. Or one can reason that if the deepened concept of voluntariness cannot be satisfied, no attribution of responsibility is ever truly justified, and all we are left with is people being coerced by their circumstances.

The pivotal question that separates the luck-tolerant cast of thought from a purist scepticism about moral agency is therefore this: what does the fact that contingency pervades human life *entail*? Adherents of both casts of thought could agree that no act is

⁴⁵ These roughly correspond to the way of the pessimist and the way of the sceptic, as Paul Russell uses these labels; see Russell (2017a).

ever purely voluntary in the way required to completely shut out contingency. It is just that while one cast of thought takes this to speak against the deepened concept of the voluntary, the other takes it to speak against the hope that praise or blame might ever be justified. It insists that blame may well never in fact be justified, but that if it *were*, this would *have* to be due to there being purely voluntary acts. The moral concepts that make up this purist cast of thought license patterns of reasoning such as the following: if anything is good, it is the moral goodness of things done from moral motives; if an action is done from a moral motive, it is a voluntary action; if an action is voluntary, it is not conditioned by anything that is contingent or lies beyond the agent's control. Via the contrapositives of those claims, one quickly gets from the realization that every action is somehow conditioned by contingent forces beyond the agent's control to the conclusion that no action is ever voluntary and moral agency has no room in this kind of world.

But one might also take the same realization that every action is conditioned by contingent forces to entail nothing of the sort. Using concepts of voluntariness, moral motivation, and moral goodness that are more tolerant of contingency and mark distinctions *within* the kind of world we inhabit, one can also endorse more luck-tolerant patterns of reasoning allowing us to accept that no moral motivation is ever fully pure of contingent inclinations, or no action ever fully pure of fortuitous influence, and still discern moral agency in the world.

To endorse the purist patterns of reasoning rather than the more luck-tolerant ones is not to fall prey to cognitive error; it is to make an error whose badness is *ethical* rather than cognitive. This reflects the needs-based approach's commitment that the authority of concepts should be assessed by looking to how they tie in with our concerns, not by how closely they approximate the set of concepts that limn moral reality. The set of moral concepts building on a deepened concept of the voluntary does not serve us well, in the sense that it does not strike a helpful balance between some of our most pressing needs and concerns. By relying on a superficial concept of the voluntary, by contrast, we can redraw the various oppositions that matter to moral reasoning *within* a world suffused with contingency, so that we can be alive to the real differences between conditioned actions that are *as* voluntary, moral, or blameworthy as actions can be and conditioned actions that are less so.

We thus need our psychological *and* our moral concepts to accommodate the fact that

our lives and our actions are suffused with contingency. As the reflection on the concerns it serves has shown, cultivating a superficial concept of the voluntary is an important step in this direction, since the concept accommodates contingency while still serving our concerns for fairness and freedom, as far as contingency will allow, by channelling attributions of responsibility away from those actions that are least within the agent's control. But it can only do this work if it resists distortion by the concern for fairness and remains superficial. Its superficiality may be a theoretical vice, but it is a practical virtue. In Nietzsche's phrase, the concept is superficial out of profundity.⁴⁶

In the next and final chapter, we return to the perspectival character of the needs-based conception of authority and explore its implications for politics. This will give us a chance to work through a second case study, in which the needs-based approach is applied to the political concept of liberty, while simultaneously rounding out our reconstruction of the Dworkin–Williams debate.

⁴⁶ See Nietzsche (2001, Preface, §4; 2005a, Epilogue, §2), who uses the phrase in a different connection. Williams (1993, 68), however, applies the remark to—among other things—the voluntary, which supports my reading on which he saw the superficiality of the concept not as a flaw, but as a feature. This contrasts with readings of Williams such as Yeager's (2006, ch. 2) and Deigh's (2008, xi).

CHAPTER NINE

The Politics of Conflicting Concerns

9.1 Three Demands of Political Disagreement

The needs-based framework presents the authority of concepts as welling up, ultimately, from the concerns of concept-users. But it is of course a basic fact of politics that different people have different concerns, and that one group's concerns can sometimes be satisfied only at the expense of another group's concerns. How does the needs-based approach accommodate the politics of pluralistic and conflicting concerns?

Far from being naively oblivious to politics, the needs-based approach comes into its own when deployed against a more politicized picture of human affairs. It equips us to address three demands arising from the fact of political disagreement: the demand that we should be able to *make sense* of such disagreements *as political* disagreements; the demand that we should be able to disagree *respectfully*; and the demand that the fact of political disagreement itself makes *on some of our political concepts*, such as the concept of liberty.

Consider first the demand that we should be able to make sense of political disagreements as political, i.e. to recognize and account for the respects in which they are political rather than epistemic in nature. One implication of the needs-based approach is that when two parties disagree, that disagreement is not necessarily a matter of epistemic error. Nor does it therefore have to be a disagreement about which concepts to use (of the kind we encountered in Chapter 1 under the headings of 'metaconceptual dispute' or 'metalinguistic negotiation'). It might rather be that the disagreement reflects an

underlying difference in conceptual needs. To hold that concepts should be tailored to conceptual needs is to adopt a standard that is local and variable enough to properly become reflective of distinctive political commitments. Some people have reason to use concepts that other people have no reason to use, or even reason not to use.

I take it to be a significant strength of the needs-based approach that it allows us to capture this political—as opposed to epistemic—dimension of disagreement. Of course, an opposition between two groups with conflicting concerns is not yet *per se* a political opposition. What makes it political is its bearing on the question of how to exercise public authority, in particular by deploying state power.¹ Unlike moral disagreement, which is characterized by the kinds of reason it draws on—to wit, *moral* reasons—*political* disagreement can draw on all kinds of reasons, and derives its political character from the fact that it is, in the end, a disagreement over how to exercise public authority. Accordingly, different cultural, social, or professional groups stand in relations of political opposition to each other precisely to the extent that their concerns conflict, i.e. are not co-satisfiable, in ways that lead them to disagree about how to exercise public authority.

On the needs-based conception of conceptual authority, it becomes intelligible for such political differences to find expression at the level of what concepts are authoritative for whom, because conceptual authority is perspectival. The concepts which are authoritative for one set of people are not necessarily authoritative for another set of people. Of course, there are limits to how much conceptual balkanization a society can sustain: the demand to tailor concepts to people's distinctive conceptual needs must be balanced against the demand, which is itself a conceptual need growing out of many of our most basic concerns, that members of the same society need to be able to communicate and cooperate with each other by *sharing* concepts. As Craig has argued in his discussion of the practical pressures towards the 'objectivization' of concepts arising in any social species such as ours, even initially fully subjectivized 'conceptual idiolects' would be driven by the demands of communication and cooperation to become usable across different perspectives and communally shared at least to some extent.² But equally,

¹ On this way of contrasting moral and political disagreement, see Williams (2005c, 77).

² See Craig (1990, 82–97; 1993, 81–115); I discuss the practical pressures to turn private thinking tools into public ones in Queloz (2021b, 145–9). See also Fricker (2010c, 61), Kusch (2011, 9–10), and Hannon (2019, ch. 2).

and especially in a pluralistic liberal democracy, there are no reasons to expect, and plenty of reasons to want to avoid, a conceptual monopoly whereby a single conceptual apparatus is uniformly shared by all and difficult to deviate from.

Secondly, as briefly hinted already in Chapter 7, such a perspectival conception of conceptual authority can yield the kind of understanding of where the other party is coming from that facilitates *respectful* disagreement. To disagree respectfully is to disagree in way of that is mindful of the requirements of respect, which notably include the idea that the other party is owed an effort at identification, i.e. an attempt to see the situation from their point of view.³

By paying attention to differences in our respective conceptual needs and what engenders them, we can perform just such an effort at identification. When the other party conceptualizes things in terms that differ from those we engagedly use ourselves, we can take up the ethnographic stance towards them: by imaginatively inhabiting and making sense of their perspective from the inside, we can come to grasp the concepts they use without making them our own. And here it can be of great value to be able to see how a difference in concepts can be rooted in a difference in conceptual needs (which would itself trace to a difference in concerns, capacities, or circumstances). For precisely what this allows us to do is to *make sense* of their perspective in a way that reveals them not to be *merely* epistemically at fault, even if their views also involve elements of such epistemic faults. A needs-based understanding of *their* reasons for concept use allows *us* to make sense of why it makes sense *for them* to use the concepts they use. Their acceptance of these concepts and their correlative reasons is not merely due to error, ignorance, immaturity, confusion, delusion, or deception. They have *reasons* to think in terms of those reasons, reasons that we can understand and rest our respect on, even if these are merely reasons for them and not for us, because we do not share the relevant conceptual needs.

As a result, we come to be able to distinguish between disagreements in which the other party is simply wrong, because they have made a mistake within a shared cognitive enterprise, and disagreements in which, though the other party may still be wrong, it is

³ Here the requirements of respect express an underlying commitment to a certain idea of equality and the ‘spirit of human understanding’ on which it is based, as Williams observed in his influential essay on equality (1973d, 236–9). I am grateful to Damian Cueni and Sanford Diehl for the pointer.

not *simply* wrong, because it makes sense *to us* that it makes sense *for them* to reason as they do, given how different from ours we understand their conceptual needs to be.

A third thing that a needs-based approach allows us to do, finally, is to ask what conceptual needs are engendered by the relations of distinctively political opposition or disagreement themselves. Even if we start from the idea that what makes for a good concept will depend on the concerns of the concept-users and accept that these vary wildly, there are concepts that serve us precisely *because* the concerns of concept-users vary wildly: concepts that are rendered needful by the fact of social pluralism. I propose to illustrate this in the next section using the example of *liberty*. We can ask how the political concept of liberty needs to be understood given a practical situation in which people have both conflicting concerns and conflicting conceptions of political values.

Perhaps counterintuitively, I want to suggest that this question need not itself be understood as a political question, but can be heard as a question in conceptual ethics. When heard in this key, the question can be answered without ending up pushing yet another sectarian creed in political opposition to its rivals. Instead, we can step back from the political fray and take it as a datum for needs-based concept appraisal: we can aim to determine whether the very fact of multiply conflictual political opposition in itself already gives us any reasons for concept use.

Some concepts may derive their authority precisely from the fact that different groups within society have conflicting concerns, because some concepts equip us better than others to accommodate such a pluralism of concerns. The very fact of living in a society marked by intersubjectively conflicting concerns will then give members of that society reason to use concepts tailored to accommodate and negotiate intersubjective conflicts of concerns. Just because they radically differ in their concerns, everyone in such a pluralistic society will have reason to use certain concepts. The conceptual authority of certain concepts will then be uniformly vindicated, not by a corresponding uniformity at the level of concerns, but, paradoxically, by a pluralism of concerns. An example of just this structure can be gleaned from the Dworkin–Williams if we only wade a little deeper into it.

9.2 The Dworkin–Williams Debate Continued

In Chapter 5, we saw Dworkin propose, in the name of conceptual integrity, to immunize

the political concepts of liberty and equality against conflict by equating liberty with *rightful* freedom. We then saw Williams resist this by suggesting, notably, that it was no good achieving conceptual integrity if it came at the cost of severing the ties to the underlying concerns that give people reason to use anything like those concepts to begin with. I used this suggestion as a springboard to a needs-based framework for concept appraisal. With this framework now in place, we can return to the Dworkin–Williams debate and use the framework to reconstruct how the argument unfolds—in particular, why the concerns animating the use of the political concept of liberty should be thought to engender a conceptual need for a concept of liberty that conflicts with the concept of equality.

Dworkin's proposed concept of liberty achieves conceptual integrity by allowing no sense in which liberty can be curtailed by rightful restrictions imposed in the name of equality. An example around which the Dworkin–Williams debate revolves is a ban on private schools imposed in the name of securing equality of opportunity in education. On Dworkin's account, if this ban on private schools reflects the will of the majority, and if that will has been rightfully implemented, it cannot have been at the expense of liberty, and no tension arises between liberty and equality.

In effect, Dworkin thus offers what Williams dubs a 'juridical' conception of liberty. It is juridical in conceptualizing liberty as extending only as far as claims in liberty are granted or denied by an agreed authority on a shared interpretative basis, because the reasons that might intelligibly ground claims in liberty are exhausted by one's rights under the political system.⁴

But political debate has far less in common with judicial review than this suggests, Williams protests: 'We and our political opponents—even our opponents in one polity, let alone those in others—are not just trying to read one text' (2005c, 78). We have different concerns and hold different outlooks articulated in terms of wildly varying conceptions that yield equally varying conclusions as to how values ought to be weighed against each other.⁵

If, in reflecting on the political value of liberty, we truly acknowledge this, and ask, as Williams congenially puts it, 'what we want that value to do for us—what we, now, need

⁴ See Williams (2005c, 86).

⁵ See Williams (2005c, 78, 86).

it to be in shaping our own institutions and practices [and] in disagreeing with those who want to shape them differently' (2005c, 75), it will emerge that what is needed in politics is not a juridical, but a more thoroughly *political* concept of liberty. A 'thoroughly political concept of liberty', for Williams, is one that 'acknowledges in its construction the on-going existence of political conflict' (2005a, 126).

These programmatic remarks clearly signal that Williams's argument lends itself to a reconstruction in needs-based terms: his guiding question is *what we, now, need the concept of liberty to be*. But his answer to that question is exegetically challenging to say the least, which helps explain why it has thus far been all but ignored even within Williams scholarship.⁶ What does it mean for a concept to be 'thoroughly political', or 'to acknowledge in its construction the on-going existence of political conflict'? And how does this amount to an argument to the effect that the concept of liberty *should* conflict with the concept of equality?

When considered within the needs-based framework, however, the pieces of the argument all neatly fall into place. Williams's starting point, we saw, is the utterly basic human concern to be unobstructed by humanly imposed coercion in doing what one wants: the pre-political concern for what he calls 'primitive freedom'. To understand how this personal concern for freedom in action relates to the political value of liberty, one must consider how such a personal concern might relate to a *political standpoint*, namely the *impersonal standpoint of society* as embodied by some public authority. From such a political standpoint, the salient question is how the often-competing concerns for primitive freedom of different people living under a shared public authority are to be registered and dealt with. To what extent should a given individual's freedom be understood as something that has any claim to society's attention?

What makes *liberty* a political concept is that it conceptualizes primitive freedom for this political standpoint: it acts as a political lens through which the concern for primitive freedom can be focused in a political context. That, at the most general level, is the role of the concept: it picks out that part of individual freedom that has a claim to society's

⁶ The only exception I am aware of is Hall (2017), who discusses the Dworkin–Williams in passing. When Damian Cueni and I set out to reconstruct the debate from a jurisprudential and philosophical perspective, respectively, we thus had to start from scratch. My indebtedness to our conversations and to Cueni's written work on this is correspondingly deep: see especially Cueni (manuscript-a, b, d).

attention.

To stand any chance of performing this role, however, the concept of liberty must differ substantially from how individuals might conceptualize their own primitive freedom. For one thing, the concept of liberty must be a *normatively richer* notion: while primitive freedom is basically a form of power that one can merely *get*, using the power one already has, liberty must be something that one can *lay a political claim to*, which is to say a claim that *makes a claim on society's attention*. The mere fact that an individual has lost some freedom does not yet give that individual any claim on society's attention.⁷ Such a claim must be backed or grounded by something other than the power one already has. This is, of course, a point that Dworkin himself acknowledges, since he likewise insists that political claims to liberty must be grounded in something (namely in rights, on Dworkin's account).

Because of its political nature, moreover, the concept of liberty must also be more *narrowly focused*: not every loss in freedom can count as a loss in liberty, as we also saw Dworkin point out against Berlin. The reason is that, as a political value, the concept of liberty has to be able to 'co-exist with the political' (2005a, 120), as Williams puts it: the concept could not intelligibly give everyone a *claim* to doing whatever they happened to want—to murder whomever they wished, for instance—because that would undermine the conditions necessary to there being any political order organized by political values in the first place.⁸ Preventing people from murdering whomever they want is a restriction on their primitive freedom, but not one that could consistently count as a restriction on their liberty, because that restriction is necessary for there to be any kind of political order at all.⁹

Someone's claim that they have incurred a loss in liberty therefore minimally needs to be *socially presentable*, as Williams put it, where that means that 'it can be urged consistently with accepting a legitimate political order for the general regulation of the society' (2005a, 120). Objecting already to the mere fact of being subject to a state at all would not be socially presentable in this sense, since it is a complaint that would apply to any state whatsoever, and the mere existence of a political order cannot consistently be

⁷ See Williams (2005a, 115).

⁸ See Williams (2001a, 93)

⁹ See Williams (2005c, 83–85; 2009, 200).

understood as constituting, already in itself, a ground for a political complaint within that order. By contrast, objecting to the operations of Franco, or James II, would be socially presentable, because ‘one could, and most objectors did, accept that these rulers should be replaced by some other rulers, and more generally they accepted a state system’ (Williams 2005a, 120).

This yields a necessary condition on a claim of a loss in liberty being correct, namely the *social presentability requirement*. But this, Williams and Dworkin agree, is not yet a sufficient condition. So what needs to be added in order to build up a conception of liberty as a genuinely political value?

Dworkin’s suggestion is that a claim in liberty needs to be *rightful*, i.e. grounded in rights. Under this conception, the reasons that can ground a claim in liberty are exhausted by one’s rights under the political system.

But on Williams’s view, this tidy picture is too tidy to make sense of the experience of life under a political order. For it is ‘one datum of that experience’, Williams stresses, ‘that people can even recognize a restriction as rightful under some political value such as equality or justice, and nevertheless regard it as a restriction on liberty’ (2005c, 84). That is to say, *even those who agree* that a ban on private schools is a rightful implementation of the majority will may still feel *resentful* of it. And it is paradigmatically through such feelings as the experience of resentment that people’s sense of freedom is given to them, and by extension also their sense of when their liberty is being restricted.¹⁰ To make sense of this experience of resentment, i.e. to conceptualize it as *reasonable* on its own terms, one needs to be able to see it as reflecting some real *loss* or *cost* in liberty that has been incurred by those who feel resentful.

But Dworkin’s proposed conception makes it impossible to conceptualize this experience of resentment as *reasonable*, because it renders unintelligible the very idea that one might incur a *cost* in liberty as a result of a rightful political decision going against one. And yet this notion of a cost in liberty incurred by those who end up on the losing

¹⁰ See Williams (2005a, 123; 2005c, 87–88); resentment is the prototypical reaction to restrictions of one’s liberty, on his account, but he acknowledges that the feelings that go with the sense that one’s liberty is being restricted do necessarily have to be identified with resentment, because ‘resentment so readily merges into other negative feelings, such as anger and dislike, not just for conceptual but also for various familiar psychological reasons’ (2005c, 87).

side of what they acknowledge to be a rightful political decision, Williams points out, ‘is at least as well entrenched in historical and contemporary experience as that of a rightful claim in liberty’ (2005c, 84). We should accordingly be suspicious of a conception of liberty that accommodates the latter but rides roughshod over the former.

However, as I propose to construe it, Williams’s objection to Dworkin’s tidy conception is not just that it is untrue to the less tidy *experience* of life under a political order. It is, more specifically, that reflection on the role of the concept of liberty ‘in political argument and political conflict’ (Williams 2005c, 84) reveals *conceptual needs* for a concept which, while narrower in scope than the concept of primitive freedom, remains wider in scope than Dworkin’s concept of liberty as rightful freedom.

In particular, the concept of liberty needs to be able to simultaneously meet the conceptual needs of those on the losing side and those on the winning side of a political decision. Those on the losing side need a concept that enables the political contestation of the majority will by enabling them to voice reasonable complaints in liberty even when these are not backed by rights. And those on the winning side need a concept that facilitates respect across the aisle by giving them the conceptual wherewithal to make sense of the costs in liberty incurred by their political opponents. Let us consider these in turn.

9.3 Enabling Political Contestation

The first conceptual need Williams invokes, on my reading, derives from the most basic concern that underlies our use of anything like the concept of liberty to begin with: the concern for primitive freedom. If the concept of liberty is to retain its connection to that concern, it needs to be able to act as the political sharp end of the concern for primitive freedom, expressing and furthering that concern in political argument.

From this perspective, having a political concept of liberty will clearly be pointless unless it allows one to lay claim to *more* freedom than one has under the current political arrangements. But the concept can only do that if it gives people grounds for complaining about restrictions on their liberty *even and especially when* they are not fully identified with the majority will and its understanding of what is rightful. A concept of liberty which *presupposes*, as Dworkin’s proposed concept does, that one is fully identified with the majority will would be pointless in this connection, since those who are fully identified

with the majority will are precisely those who have no cause for complaint.

To serve the concern for freedom in the political sphere, we therefore need a concept of liberty under which the reasons grounding claims in liberty are not exhausted by one's rights, but allow for reasonable complaints in liberty grounded in the residual losses in liberty one has incurred as a result of the rightful implementation of the majority will. The concept of liberty can meet our needs as a political concept only if it gives us grounds for *contesting* the prevailing understanding of rightful freedom. And, to enable such contestations, the concept of liberty needs to make conceptual room for the thought that a rightful restriction on someone's freedom can still *reasonably* be resented as a *loss in liberty*. For only then can the concept of liberty serve our concern for freedom in such situations—and such situations are sure to arise as long as there are people who are not fully identified with the *volonté générale*.

Dworkin's conception of liberty as rightful freedom, by contrast, leaves no conceptual room for reasonable complaints in liberty from those who end up on the losing side of a political decision. They can of course still complain—but their complaints must appear confused in light of this conception, for, under this conception, the reasons that can ground their claims in liberty are exhausted by their rights under the political system.

Any construction of the concept of liberty that is to enable political contestations of what is to count as rightful freedom therefore needs to spread the idea of liberty, and hence of a cost in liberty, more widely than Dworkin's proposed concept does. While the concept of liberty cannot, consistently with the very existence of the political, treat any and all complaints in primitive freedom as reasonable—‘no concept of liberty intelligible as a political value could allow anybody to murder anybody they liked’ (Williams 2001a, 93)—a helpful concept of liberty, and particularly one that is suitably responsive to the fact that pluralistic societies are unlikely to be of one mind about everything, will need to be far more inclusive than Dworkin's if it is to serve the concern for primitive freedom by facilitating the political contestation of the majority will. We will still want to rule out as unreasonable claims that are not socially presentable in Williams's technical sense; and we will also want to rule out claims that are merely a product of insincerity or of insufficient information or attention to the relevant arguments. But when these fairly minimal demands are met by someone's claim that they have incurred a loss in liberty, then, though we may not yet have a *rightful* claim, we will have what Williams calls a

responsible claim.¹¹ And, for political purposes, we should regard the voicing of a responsible claim of a loss in liberty as a sufficient reason to assume that there *has been* a loss in liberty.

This is more than the blank assertion of a liberal piety. The conceptual need engendered by the concern for freedom constitutes a good second-order reason to prefer a conception of liberty along these lines to the one advocated by Dworkin. In leaving no conceptual room for complaints, on the part of those whose desires are frustrated by political decisions, that they have incurred a cost in liberty, Dworkin forecloses an important form of political argument—a form of argument which, especially in pluralistic societies in which full and general identification with the *volonté générale* is hard to come by, is crucial to the political expression of the concern for freedom.

A concept of liberty capable of serving the concern for freedom of those who end up on the losing side of a rightful political decision thus needs to allow for reasonable complaints in liberty grounded in the residual losses incurred by this decision. In other words, Dworkin and Williams agree that not every loss in primitive freedom can be reasonably resented as a loss in liberty, since the concept of liberty, just because it is a political concept, needs to be narrower in scope than the concept of primitive freedom; but Williams insists against Dworkin that not every *reasonably* resented loss in liberty has to be *rightfully* resented as such, because the political concept of liberty needs to be broader in scope than the concept of rightful freedom.

Of course, this will cast the net for complaints in liberty that are intelligible input to political debate fairly wide, but if the concept of liberty is to serve the concern for freedom, the net *needs* to be cast wide, because how seriously any of these complaints should be taken is itself a political question, not one to be settled in advance by a

¹¹ Williams writes: ‘a claim to the effect that one’s liberty is being infringed is *responsible* if the claim-maker makes it sincerely, is convinced that the political order is legitimate, and moreover remains persuaded of his claim despite attending to serious argument, and so forth. ... The suggestion I want to make is essentially that ... a responsible claim to a loss of liberty means that there is a loss of liberty’ (2005a, 122). The shape and language of Williams’s suggestion here leaves it in no danger of being mistaken for a metaphysical argument resting its weight on an independent account of the intrinsic nature of liberty; it is clearly an argument in conceptual ethics, which advocates a certain way of thinking in politics: namely, one on which we *count* responsible claims of a loss in liberty as *reason enough* to assume that there has been a loss liberty.

definition. The concept's role, insofar as it serves the concern for freedom, is to determine what can go into the funnel of political debate, not what comes out of it. As Williams puts it:

A construction of liberty on these lines ... means that, within certain limits, anyone with a grievance or who is frustrated by others' actions can appropriately complain about restrictions on his liberty. If 'appropriately' means that it is semantically, conceptually, indeed psychologically, intelligible that he should do so, that is right. If it means that it is necessarily useful, helpful, to be taken seriously as a contribution to political debate, and not a waste of everyone's time, it is not right. The point is that these latter considerations are in the broadest sense political considerations, and that is the point of the construction. (2005c, 92)

This conception of liberty determines what losses in primitive freedom merit to be so much as *intelligible* as input to political debate, but it does not determine what should come out of the debate.

Dworkin's conception, by contrast, collapses these two steps into one, and, in so doing, entrains a loss in freedom: in particular, a loss in the freedom to contest rightful restrictions of liberty through complaints that have a claim on society's attention.¹² The concept of liberty can only serve the concern for freedom in this regard if it casts liberty as broader than rightful freedom. Instead of serving the concern for freedom that most basically animates the use of the concept of liberty, Dworkin's conception of liberty thus ends up working against the concern for freedom.

At its core, then, Williams's first objection is an argument from the concern with primitive freedom to what an adequate political concept of liberty needs to be. The concept of liberty fundamentally serves the concern with primitive freedom, focusing that concern in the political context. Dworkin's proposed conception fails to serve that concern in one crucial respect: it achieves congruence with the concept of equality at the cost of failing to serve people's concern for freedom when their freedom is rightfully restricted. A concept of liberty able to serve people's concern for freedom under those

¹² The focus here is on claims in liberty in political debates, but parallel considerations apply also to the *legal* treatment of complaints about restrictions on freedom. Damian Cueni spells this out in an article arguing that due respect can be paid to reasonable claims in liberty even when they are not ultimately found to be rightful thanks to the legal distinction between rights infringements and rights violations; see Cueni (manuscript-d).

circumstances cannot therefore coincide with the concept of rightful freedom.

9.4 Respect Across the Aisle

The second conceptual need that Williams takes a thoroughly political concept of liberty to answer to is the need to support the containment, within a shared political system, of pluralistic and conflicting concerns. We might put this by saying that the pluralism of concerns itself engenders a conceptual need for a concept of liberty that facilitates our living together in a pluralistic society with others whose concerns radically differ from our own. One notable way in which a concept of liberty can facilitate this is by enabling *respect across the aisle*—in particular, by equipping those on the winning side of a political decision to make sense of and acknowledge the costs in liberty incurred by those on the losing side instead of treating their complaints as products of error, ignorance, immaturity, confusion, delusion, or deception. It makes a great difference to our ability to live together whether we have the conceptual wherewithal to make sense of the resentment of our political opponents as reasonable.

What does Dworkin's tidy equation of liberty with rightful freedom—the freedom that each has a claim to, based on equally distributed rights—allow him to *say to* those who sharply feel that they have incurred a cost in liberty even though their liberty has been rightfully curtailed? All that Dworkin can say to them is that they are mistaken: they may *think* they incurred a cost in liberty, but if only they achieved a clear-headed understanding of the concept of liberty, they would recognize that no such loss was incurred, and their resentment was therefore unreasonable.¹³

Against this, Williams points out that '[t]elling these people that they had better wise up and revise their definition of the values involved' as Dworkin's proposed conception invites us to do, 'is not in many cases prudent, or citizenly, or respectful of their experience' (2001a, 102). In effect, Williams here extracts from a pluralist understanding of politics a sense of the conceptual needs to which the concept of liberty must answer—a sense of what we need the concept of liberty to be. 'The idea of value pluralism', he

¹³ What Dworkin tells the people who feel they have incurred a loss in liberty, Williams concludes, is thus essentially what Rousseau once told them: 'that they did not really understand what liberty was' (2001a, 100). See also Williams (2005c, 85).

writes, is no mere ‘aestheticism of politics’; it ‘tells you how to speak to the people who have to pay, not just in their interests but in their values, for things that have to be done’ (2001a, 102). It would be neither respectful nor prudent to dismiss their complaints as conceptually confused. The ability to secure the consent of those who end up on the losing side of a political decision is crucial to the stability and health of a democracy.¹⁴ And whether that consent can be secured depends in no small part on *what we can say* to those who feel they have incurred a cost in liberty as a result of a decision going against them.

As the legal scholar Jamal Greene remarks, the last century may have given us tools ‘to fight political exclusion’, but ‘in this century, we need the tools to build a politics of pluralism’ (2021, xxi). The concept of liberty that Williams advocates is just such a tool: it is tailored to the conceptual needs of the politics of pluralism.

This is where we find an argument that attempts to demonstrate the society-wide authority of a certain conception of liberty not from equally widely shared concerns, but, on the contrary, from the conceptual needs arising precisely from societal variance in people’s concerns and conceptions of political values such as equality. Just because pluralistic societies are not of one mind about what political values such as equality require and how they should be weighed against other values, Williams suggests, we need a conception of liberty that allows us to make sense of the complaints of those who reasonably resent what is being enacted in the name of such things as equality:

Even though you and I share a certain conception of equality, and are happy to see its policies being enacted, we can and should use a political concept of liberty in terms of which we can not only sympathize but agree with our fellow citizen who does not share this conception of equality, resents what is being done to him in its name, and says that he has lost some of his liberty. He is reasonably resentful of what is happening to him, because he is being coerced against his will; and these are the complaints of someone who accepts, not only some political order, but this one. Can we really tell him that if he only understood liberty better, he would see that he was deluded in thinking that he had lost some? (2005a, 126)

Adopting this attitude towards our political opponents’ complaints, Williams maintains

¹⁴ On losers’ consent and its importance to democratic legitimacy and stability, see the essays in Anderson et al. (2005).

against Dworkin, is ‘objectionable’ (2005c, 85). In particular, it is not respectful of our political opponents’ experience. We should ‘take seriously the idea that if, under certain conditions, people think that there is a cost in liberty, then there is’—this is a condition ‘not only of taking seriously the idea of political opposition, but of taking our political opponents themselves seriously’ (2005c, 85).

To this end, we need a concept that is less ‘instructional’ and ‘patronizing’ than Dworkin’s concept.¹⁵ We need a concept that enables those on the winning side of a political decision to *respect* those on the losing side *as* people who have suffered costs in liberty, which in turns means that we need a concept of liberty that makes conceptual room for the thought that even a *rightful* restriction on someone’s liberty can still *reasonably* be resented as a loss in liberty. Having the conceptual wherewithal to acknowledge the costs they suffered as a result of a rightful decision, and to treat them respectfully rather than patronizingly, is part of what is required to foster respect across political divides, and thus ultimately also to secure losers’ consent. Hence Williams’s emphatic conclusion that ‘*the proposed interpretation of liberty is what we need in order to live in society with others who have different interpretations of equality*’ (2005a, 125–26). We need this concept of liberty because it facilitates and fits a respectful attitude between political opponents. It enables respect across the aisle, and thereby, in Chantal Mouffe’s contrast, prevents adversaries from turning into enemies.¹⁶

The relation of political opposition is a relation that not only specially calls for respect, but also leaves more space for it.¹⁷ After all, as Williams repeatedly emphasizes, a political decision ‘does not in itself announce that the other party was morally wrong or, indeed, wrong at all. What it immediately announces is that *they have lost*’ (2005j, 13). That is why regarding someone as a *political* opponent is different from regarding them as someone one is *morally* at odds with. In a moral disagreement, we treat the other party as people we simply have to argue with to make them see that they are morally wrong. But in a political disagreement, Williams thinks, ‘[w]e should not think that what we have to do is simply to argue with those who disagree: treating them as *opponents* can, oddly

¹⁵ See Williams (2005c, 86).

¹⁶ See Mouffe (2000, 13).

¹⁷ On the idea that political opposition specially calls for respect, and that legitimate opposition is an achievement worth defending, see Kirshner (2022).

enough, show more respect for them as political actors' (2005j, 13). In particular, it leaves more room for the kind of respect that Stephen Darwall calls 'recognition respect', in contrast to 'appraisal respect'.¹⁸ Whereas appraisal respect consists in respecting people for their display of some form of excellence, recognition respect consists in giving appropriate consideration or recognition to some feature of other people in one's practical deliberation. In the relation of political opposition, this feature, which we express respect by duly recognizing, is, most fundamentally, the fact that they disagree with us, and that they do not necessarily do so merely out of error. We recognize the possibility that they might have come by their political convictions much as we came by ours, namely through an obscure confluence of sociohistorical and biographical circumstances, passions, and interests. On that basis, we can then recognize that a political decision going against them may come at a real *cost in political value* to them. This is a precondition—and it is no more than that—of acknowledging what resentment this produces as reasonable, and of meeting one's political opponents halfway.

Such respect across the aisle requires the necessary concepts, however, because respect is characteristically respect *under a description*: we do not just show respect *for* people, but for people *as* people who possess some property, and this requires the conceptual wherewithal to make sense of them as possessing that property. Even for it to make sense to show respect for persons *as persons* already requires not just some concept of a person, but, specifically, a conception on which the fact of being a person itself already entitles one to respect under that description. A fortiori, for it to make sense to show respect for persons *as persons who have incurred a cost in liberty when their liberty has been rightfully curtailed* requires a conception of liberty that allows at least some claims in liberty to remain reasonable beyond the point at which they cease to be rightful.

If a certain conception of liberty can help us to make sense of our political opponents' resentment, this is also due to our ability to use concepts in a disengaged way. Besides thinking in terms of the conceptions of political values that we are engaged users of, we also have to be capable of thinking, if only in a disengaged way, in terms of the different conceptions of political values that 'move around society in their variously resentful or hopeful ways' (Williams 2005a, 126). Recognizing that some might reasonably resent the enactment of some policy we are fully identified with ourselves, given how we

¹⁸ See Darwall (1977).

conceive of our own political values, requires the ability to imaginatively identify with the perspective of those who view things in terms of rival conceptions, and recognize how these conceptions would give them grounds for resentment. This feat of disengaged or vicarious resentment requires what Williams calls ‘double-mindedness’ (2005a, 125–26). By making disengaged use of their conceptions even while remaining fully engaged users of our own, we can come to discriminate between restrictions they could or could not *reasonably* resent.

As the qualification ‘reasonably’ registers, however, it makes a difference whether their conceptions really are *authoritative for them*, i.e. whether, in light of their concerns, capacities, and circumstances, they in fact have reasons to think in these terms, or feel resentful merely because they are in the grip of conceptions that they would immediately jettison if only they reflected on how these related to their concerns.

Whether it makes sense *to us* that it should make sense *for them* to think in these terms will thus also inform our judgement as to whether they have a solid basis for reasonable resentment. If only inchoately, our sense of our political opponents’ reasons for concept use plays a part in determining how seriously we take the reactive attitudes they form in light of those concepts: if we feel that their resentment is entirely an artefact of their confidence in notions that seems misplaced even by their own lights, in that they would themselves reject those notions upon reflection, we will find it difficult to view their resentment as reasonable; conversely, if we feel that they have good reasons to think as they do, they will seem to us to be on solid ground, and their resentment will at least seem reasonable on their own terms, even if these are terms we do not share. Thus, whether claims in liberty are *responsible* depends not just on whether they are made sincerely and after due consideration, but also on whether they are made on the basis of *responsible conceptions* of political values.

When ‘double-minded’ reflection on our political opponents’ first- and second-order reasons for thinking as they do reveals their conceptions of political values as well as the claims they make on that basis to be responsible, this licenses respect across the aisle: it enables us to take a respectful view of those political opponents as standing on ground as solid as our own, in that they appeal to conceptions that are as authoritative for them as ours are for us. Consequently, when they end up on the losing side of a political decision, it enables us to take their complaints seriously. It makes sense to us that some other way

of viewing things should make sense to them. We may not agree with them, but we can also see that they are not simply conceptually confused.

The Dworkin–Williams debate thus boils down to the following disagreement: for Dworkin, we have a standing reason to prefer concepts or conceptions that do not conflict, and this gives us a reason to use a concept of liberty under which the reasons grounding claims in liberty are exhausted by one's rights under the political system. For Williams, on the other hand, the pursuit of conceptual integrity must take a backseat when it threatens to sever the connection to the underlying concern for freedom that most basically animates our use of anything like the concept of liberty, and this is what Dworkin's proposed concept threatens to do when it denies people any reasonable claims in liberty beyond those grounded in rights.¹⁹ In the process of whittling away the aspects of the concept of liberty that bring it into tension with the concept of equality, moreover, Dworkin also ends up shaving off a second valuable feature of the concept of liberty: that it facilitates respectful political disagreement by enabling those on the losing side of a political decision to make claims in liberty that those on the winning side can acknowledge as reasonable even when they are not backed by rights.

On this needs-based reconstruction of Williams's argument, there are therefore two reasons to prefer a conception of liberty along the lines that Williams suggests over Dworkin's tidier conception. First, to serve the concern for freedom in the political sphere, we need a conception of liberty under which the reasons grounding claims in liberty are not exhausted by one's rights, but allow for reasonable complaints in liberty grounded in the residual losses in liberty one has incurred as a result of the rightful implementation of the majority will. This allows those on the losing side of a political decision to contest the majority will. Second, to facilitate respect across the aisle, we need

¹⁹ Note that Williams's point is not that Dworkin's preferred concept of liberty is inadequate because it fails to serve the *concern for liberty*. Since the concern for liberty is focused by the concept of liberty, arguing for a particular understanding of the concept of liberty based on the concern for liberty would be circular, a form of self-validation that precisely fails to leverage any common ground with those who, like Dworkin, favour a different conception of liberty. Williams agrees that—in Dworkin's phrase—'part of politics consists in arguing about' (2001b, 255) what precisely political concepts like liberty and equality amount to, and hence what the concern for liberty and the concern for equality should be concerns *for*. Yet Williams accommodates this fact better than Dworkin does, because he argues *from* that fact *to* a certain conception of liberty.

a conception that equips those on the winning side of a political decision to make sense of and acknowledge the costs in liberty incurred by those on the losing side. This need can also only be met if the conception does not limit reasonable claims in liberty to those that can be grounded in rights. For both reasons, therefore, the concept of liberty we need for the purposes of political debate had better not understand all political claims in liberty in terms of rights—which in turn means that it had better not be immunized against conflict in the way that Dworkin proposes.²⁰

For the needs-based approach to conceptual authority, the broader lesson is that even a highly conflictual pluralism of concerns need not be fatal to the prospects of uniformly shared reasons for concept use. Williams's argument interestingly draws uniformly shared reasons for concept use out of the observation that politics must contend with differences and conflicts of concerns. Under circumstances of liberal democratic pluralism, only a concept of liberty that is more elastic than the concept of rightful freedom will allow us to understand and respect our political opponents when, in response to some measure enacted because they lost at the ballot box, they insist that they incurred a loss in liberty. The very fact that people's concerns are at variance can therefore itself give us certain reasons for concept use.

9.5 Placing the Demand for Theoretical Virtues

Finally, however, we can also draw on the needs-based framework to come to Dworkin's defence, and find a place for his tidy-minded insistence on the value of concepts exhibiting theoretical virtues. If our concerns are such that we sometimes need incongruent and tension-ridden concepts, then they might equally be such that we sometimes need tidy and congruous concepts. In other words: if we can only *place* Dworkin's pursuit of theoretical virtue within some need matrix in which it answers to genuine conceptual needs, that pursuit will to that extent be vindicated even on the needs-based approach.

Once we buy into the idea that tidy-mindedness must be answerable to a *need for*

²⁰ This insistence that not all political claims in liberty should be understood in terms of rights echoes a complaint about Dworkin's approach that Williams voiced already in 1986, namely that 'there is something perverse in the demand to force all principled political argument into this one mould, and to ignore the wider range of conceptions that certainly have power in our political discourse' (2014b, 260).

tidiness, the question ‘Should we tidy up our concepts?’ assumes a markedly different shape. In particular, it no longer looks like a binary yes-or-no question, because we can have *more or less* of a need for tidy, theoretically virtuous concepts; and it no longer appears to be answerable in the absolute, because the extent to which we have a need for tidy concepts can only be determined in relation to a concrete practical context in which the need arises out of a combination of concerns, capacities, and circumstances. Even if it can be shown that tidy-mindedness answers to a real need, therefore, it does not follow that there always is such a need. Rather, the need for theoretically virtuous concepts will be *scalable* and *context-sensitive*—a need that grows out of, and varies with, concept-users’ practical situations.

In fact, even Williams, though known as an arch-critic of theory-construction in ethics, grants that once we reflect on when and why people might experience a need for more systematic concepts, a legitimate place *can* be found for the aspiration to reduce conceptual conflict by tidying up thought:

If... conflict is not a logical affliction of our thought, it must be a mistake to regard a need to eliminate conflict as a purely rational demand, of the kind that applies to a theoretical system. Rather we should see such needs as there are to reduce conflict and to rationalise our moral thought as having a more social and personal basis. In particular, in a modern complex society functions which are ethically significant are performed by public agencies and, if the society is relatively open, this requires that they be governed by an explicable order which allows those agencies to be answerable. (1981a, 81)

The key idea hinted at here is that tidy-mindedness answers to a need to *hold public decision-making accountable*. Williams never quite spells out this line of thought, but by piecing together various remarks and elaborating them into a continuous line of argument, we can narrow the chasm between Williams and Dworkin, and illustrate how the needs-based approach can accommodate the Dworkinian aspiration to tidy up thought, by underscoring how the conceptual needs of people in the position of public authority are quite different from those of people in the position of citizens engaged in political disagreement with their fellow citizens.²¹

²¹ I draw here on an article I co-authored with Damian Cueni on the practical pressures that give rise to the demand to explicate values in terms of stateable principles, to make them consistent by using some of them

The place to start is with conflicts between people rather than concepts. Whatever exactly their concerns are directed to, people will come into conflict with each other as they seek to meet those concerns: the exercise of one person's primitive freedom is bound at some point to get in the way of another person's primitive freedom. Such restrictions on one's freedom by other intentional agents rather than by the blind forces of nature are particularly prone to breed resentment: as Rousseau notes, we 'endure patiently the necessity of things but not the ill will of others' (1979, 91)—one may stoically accept being locked in a hut by a blizzard, yet grow deeply resentful if locked in by someone else, even though the degree of restriction on one's freedom is the same.²² To prevent the antisocial sentiment of resentment from dissolving social relations, some sort of impartial or public authority is needed that can resolve these conflicts, determining the priority of conflicting parties' concerns and how far these are to be met.

To be effective, however, that impartial or public authority will need the power to enforce its decisions, and whenever this power to enforce is used coercively, that coercion again risks arousing resentment. Coercion by purely arbitrary assertions of will—even if it is the will of an impartial authority—is just as much coercion as coercion by other parties. To avoid recreating, at a different level, the very problem of coercion-induced resentment that it was meant to solve, 'the authority needs to *have* authority' (2005c, 94), as Williams puts it: the people on whom the decisions are enforced need to be able to differentiate the enforcement from naked coercion by making sense of the decision-making as authoritative. It is the authority of impartial conflict resolvers that distinguishes their decisions from coercion and allows them to quell rather than breed resentment.

If we now consider how this highly general and underdetermined schema has been elaborated in the modern world, the most significant fact, as Williams points out with a nod to Max Weber's distinction between rational-legal, traditional, and charismatic authority, is that the modern state is 'a formation in which authority is peculiarly vested

to overturn others in systematic ways, and to render them as far as possible discursively justifiable; see Cueni and Queloz (2021). I am also indebted to Cueni's (manuscript-a) account of how Dworkin's position is more charitably construed as reflecting a pressure on the judiciary to speak with one voice.

²² See Geuss (2001, 104–5).

in discursive argument, rather than in traditional or charismatic leadership' (1996, 33).²³ This partly reflects the fact that many of the traditional stories that formerly enabled societies to make sense of decision-making as authoritative—stories about divine right, revealed knowledge, or natural hierarchies, for instance—no longer carry enough conviction in the modern world; and it partly reflects the sheer size of modern societies, which makes it much harder for personal trust in public decision-makers to be sustained on the required scale: insofar as one is personally acquainted with a public decision-maker, one might trust them not to decide arbitrarily, which would leave them free to resolve conflicts by exercising their judgement and drawing on whatever concepts have force with them, much as people do in their private deliberation. In large-scale societies, however, not everyone can be personally acquainted with every public decision-maker,²⁴ and those subject to the decisions will consequently want to be able to *ascertain* that the decision-making was an exercise of reasonable judgement rather than an arbitrary assertion of will.

In large modern societies, and especially in liberal democracies, the concern for primitive freedom thus gives rise to a need to hold public decision-makers to account by *demanding discursive justifications* for their decisions (as opposed to simply trusting them to make the right decisions). Yet this alone still provides little check on arbitrariness if today's declared justifications are permitted to be entirely inconsistent with yesterday's. This helps explain why 'there is a demand of rational consistency and principle in public positions' (Williams 2006h, 164). Non-arbitrariness requires *consistency over time*.

In their private deliberation, individuals may be allowed to rely on intuitive and particularized judgement that draw on conflicting concepts and adjudicate conflicts on a case-by-case basis, without formulating explicit and general principles by which to justify each decision. But the demand for accountability drives public decision-making away

²³ Weber (2019, 338–78) distinguishes three *Idealtypen* or pure types of authoritative rule or leadership which in reality are often combined in varying proportions: the rational (or rational-legal) type derives its authority from formal rules and a legally formulated impersonal order; the traditional type derives it from the sanctity of established traditions and customs; and the charismatic type derives it from the special personal qualities or powers of a leader. On Weber's influence on post-World War II Oxford, see Finnis (1985).

²⁴ On the first point, see Williams (2003, 117–8; 2005c, 95–96; 2009, 200–1); on the second, see Cueni and Queloz (2021).

from this intuitive and unsystematic condition. To be *verifiably* non-arbitrary, decision-making needs to be discursively justified by more general and explicitly stateable principles, which moreover need to exhibit some degree of systematicity because they need to be consistent with a series of past decisions and their discursive justifications. Intuitive and unsystematic judgements whose authority rests on personal charisma or tradition thus come under pressure to become verifiably non-arbitrary by assuming a more impersonal, principled, and systematic form.

The demand to give reasons for reasons has an important place in these dynamics, since one way in which public decision-making will have to be answerable to the public is by having something to say about *why* the reasons that guide public decision-making should be regarded *as* reasons. As Williams observes,

some distinction, not further reasoned, can ground agreement in private and less impersonal connections, but may not serve, or may not continue to serve, where a public order demands a public answer. To take an example which has been recently discussed, a distinction between abortion, which is permitted, and infanticide, which is not, is one which can probably be naturally sustained in a certain context of shared moral sentiment without further reason being needed. The fact that further reason is not needed does not mean that that distinction is *irrational*. It means only that the basic distinction is more directly convincing than any reason that might be advanced for it: another way of putting it is that ‘You can’t kill that, it’s a child’ is more convincing as a reason than any reason which might be advanced for its being a reason. (1981a, 81)

As long as confidence in the concept providing the reason is sufficiently strong and communally shared, the use of the concept and its correlative reasons need not be vindicated by a further reason; for those who are fully confident in the concept, the reasons it immediately provides will be more compelling than any reason that might be offered in support of reasoning in those terms, just as the reason not to kill provided by the applicability of the concept *child* plausibly counts as stronger—for anyone who confidently deploys the concept—than any reason that could be offered in support of treating something’s being a child as a reason not to kill it. The reasons *in* concept use are felt to be so convincing that reasons *for* concept use are not required. So long as this confidence in the concept is both strong and shared between the decision-makers and those to whom the decision must be justified, no reasons for those reasons will be

demanded.

But the larger and the more pluralistic societies become, the less it can be taken for granted that such confidence, especially confidence in highly variable thick normative concepts, will in fact be broadly shared. An increase in society's size and pluralism is likely to bring two developments in its wake.

On the one hand, it is likely to entrain an increasing need for *further rationalization* of the discursive justifications offered: the fewer the concepts that can be assumed to be shared between those doing the justifying and those to whom the justifications are addressed, the more it will be felt that the reason-providing concepts at work in the justifications are themselves in need of reasons bolstering their authority. That is one important way in which the authority question comes to play a role in modern life. Public decision-making is subjected to a demand for further rationalization that private deliberation is not subjected to in the same way.

On the other hand, greater pluralism encourages a *shift from thicker to thinner* concepts in public discourse: just because thick concepts are richly expressive of a certain social and evaluative perspective, they are only forceful as long as that perspective is shared between justifiers and their addressees, and the greater the variety of perspectives that a society harbours, the more the stock of unquestioningly shared thick concepts shrinks. There will then be a corresponding conceptual need for public justifications to abstract away from the peculiarities of distinctive perspectives and to draw on thinner concepts, arrived at through reflection on what remains common to different perspectives articulated in terms of different thick concepts.²⁵ Under the pressure to offer a widely intelligible justification, public decision-makers are driven to retreat from the more idiosyncratic components of their conceptual apparatus and fall back instead on concepts that are more widely shared among their addressees (for instance, because they are more *formal* or *procedural* concepts). This retreat to more widely shared concepts under the pressure to justify one's judgement to a wider audience is not limited to ethical or political concepts. A radiologist looking at a CT scan can distinguish the radiological features of brittle bone disease from the radiological features of child abuse by deploying concepts that are the preserve of specialists; but when pressed to justify that judgement in terms

²⁵ For an exploration of the possibility of thin normative concepts with different extensions and its implications for metaethical realism, see Eklund (2017).

that make sense to non-specialists, she will retreat to concepts that are more widely shared, such as *line*, *angle*, *thinning*, or *mosaic pattern*. When the concepts are ethical or political, such a retreat to more widely shared concepts *may* contribute to shaking off parochial prejudices and biases encoded in thicker concepts; but there is no guarantee that it will do so—there are notoriously many ways in which such biases can be reintroduced even into thinner, more widely shared conceptual frameworks.²⁶

Of course, the need for further rationalization does not necessarily *entail* a shift from thick to thin—a highly homogeneous community might conceivably share a large stock of thick concepts that allowed that need to be met by drawing on further thick concepts. In a less homogeneous society, however, the need for the further rationalization of discursive justifications *also* drives the thinning out of thought. In particular, the shift to thinner concepts becomes inevitable in public contexts whenever they are the only concepts that allow one to articulate, in terms intelligible to all relevant parties, the consensus obtainable at the intersection of different perspectives.

Within the needs-based framework, the conceptual need for discursive justifications to be further rationalized in terms of reasons articulated in thinner terms can be seen, then, not as a ‘demand of pure rationality’, but rather as a need growing out of ‘a certain kind of public order’ (Williams 1981a, 81). It grows out of the concern to honour a certain ideal of public life in liberal democracies, an ideal that ‘requires in principle every decision to be based on grounds that can be discursively explained’ (Williams 2011, 20). If public agencies are to be answerable to the *demos*, their decision-making needs to be governed by an explicitly justifiable order that makes sense from a plurality of perspectives.

This need paradigmatically applies to the state’s communication with its subjects, but it is not confined to it. It applies equally to decision-making bodies in private institutions,

²⁶ Post-colonial critiques of the ostensibly universalized concepts of international law offer examples of this. See, e.g., Pahuja (2011), which offers a darker retelling of the process of decolonization, the establishment of Permanent Sovereignty over Natural Resources, and the assertion of a rule of international law after the end of the Cold War; in all three instances, she argues, the promise of betterment through the universalization of legal doctrines such as sovereignty and self-determination masked attempts to reproduce, through economic arrangements, the colonial power relations that used to be explicitly articulated in the political sphere. See Cueni and Queloz (2022) for a discussion of this and related examples.

including hospitals and universities. This reflects another characteristic development of modernity, namely that ‘the extent of the public is growing’, as Williams puts it: issues of ethical and political significance are increasingly ‘governed by regulations that are publicly declared and debated’ (2005h, 45). Difficult decisions that used to be left to the judgement of private individuals—think of doctors, who, especially before the rise of the modern hospital, visited patients in their homes, and made whatever hard calls they had to make all by themselves, *in foro interno* or at least within the privacy of those homes—are increasingly being taken within more institutionalized settings of public concern, and are therefore appropriately shouldered by decision-making bodies such as hospital ethics committees. Or, if they do not themselves take the decisions, public commissions, ethics committees, or government panels are increasingly tasked to regulate how such decisions are taken.²⁷

This affects both the way in which morally or socially significant issues are addressed and our understanding of what counts as a rational way of addressing them. When thorny issues are settled through silent personal deliberation on a case-by-case basis, there is a great deal *less* pressure to think in explicit, determinate, principled, consistent, generally applicable, and widely shared terms than when such issues are adjudicated through formalized committee meetings and public regulations. That is not to say that silent personal deliberation is devoid of any pressure in that direction: careful attention to its workings and conceptual needs is likely to uncover a demand for some degree of generalizing and principled thought even there.²⁸ But there is certainly a great deal *more* pressure to recast one’s thought in more theoretically virtuous terms in the context of public deliberation. It is, above all, the conceptual needs of high-quality, widely accessible, transparent, and accountable public deliberation that require the implicit to be made explicit, the vague determinate, the unprincipled principled, the inconsistent

²⁷ See the valuable discussion of this development in Harcourt (Manuscript), who connects it to reflections by Williams (2005h) and to a discussion on ‘the administered world’ between Theodor Adorno, Max Horkheimer, and Eugen Kogon (1989).

²⁸ Attention to the distinctive conceptual needs to which different forms of thought are answerable should thus not make us too quick to follow radical particularists like Jonathan Dancy (2004, 2017) in excising principles from personal deliberation; see Heney (2016, 134–38) for a critique of Dancy along these lines.

consistent, and the parochial general.²⁹ If the extent of ‘the public’ is growing, therefore, then so is the pressure on practical thought to approximate a tidy system.

But just how much like a tidy system does thought need to become, on this view? If we understand the demand to systematize—i.e. to rearticulate thought in more theoretically virtuous terms—not, in the first instance, as an unconditional and universal dictate of rationality itself, but as a conceptual need growing out of a concern to render public decision-making accountable, this transforms the binary question ‘To systematize or not to systematize?’ into a question whose answer varies from one context to another and with regard to the degree of systematization required. Conceptual needs are not standing imperatives that apply regardless of context, and they are not binary, flicked on or off according to whether their needfulness conditions are satisfied or not. They must be thought of in more dynamic terms, as functions of needfulness conditions that are fulfilled to varying degrees in different contexts. In short, conceptual needs are context-sensitive and scalable.

We can derive a dynamic understanding of the need to systematize thought from a *parametric* need matrix that presents the resultant need as a function of the following three parameters *within* its needfulness conditions:³⁰

Three Parameters of the Need to Systematize Thought:

[Who] needs to give reasons for [what] in terms that must make sense to [whom]?

The first parameter concerns *who* is subject to the need: is it everyone capable of practical reason, for instance, or merely people in positions of public authority? The second parameter concerns *what* those subject to the need must be able to give reasons for. Do they have to justify any conceivable case in a manner consistent with any other conceivable case, or only the cases that actually arise within a certain society in a manner consistent with the past cases that actually arose in that society? The third parameter concerns *to whom* the discursively articulated reasons for the decisions are addressed and need to be intelligible. Do these reasons have to make sense to anyone capable of practical

²⁹ Some of the pressures in this direction are emphasized by O’Neill (1987), Page (1996), Zacka (2017), and Prescott-Couch (2021).

³⁰ In this and the following four paragraphs, I draw, with various modifications, on Cueni and Queloz (2021).

reason, or can they draw on the shared conceptual resources and concerns of a more local constituency? Of course, the reasons must make sense of the decision to those addressees not just in the minimal sense of rendering it *humanly intelligible* ('It's just what bandits do'), but in the normative sense of presenting it as *authoritative* for them.

In light of this parametric need matrix, the applicability, scope, and strength of the need to systematize changes with the values of these three parameters. Consider what form that need would take if all three parameters were dialled up to their maximum. This would mean that *everyone* capable of practical reason would need to be able to decide and justify *any conceivable case* in a manner consistent with *every other conceivable case*, and in terms that made sense to *anyone* capable of practical reason. The need for systematicity would then be ubiquitous, and it would take a very demanding form, pushing concepts-users all the way to a highly systematic and thinned out conceptual apparatus: it would have to be highly systematic in order to yield reasons applicable to and consistent with any conceivable case; and nothing short of highly general considerations and principles articulated in terms of very thin concepts would do, since any more thickly perspectival considerations would be disqualified as too contingent and counterfactually fragile to yield justifications that made sense to anyone capable of practical reason. We can already discern, in this demand that everyone be able to decide and justify any case in terms that make sense to anyone, something of the universalist and rationalist spirit that drives the development of theories such as Kantianism or utilitarianism.

If we understand the need for systematicity as growing out of a concern to render public decision-making bodies accountable to those whom the decisions affect, by contrast, this suggests a more specific way to set the three parameters. It suggests, first, that those who need to systematize are *people in the position of public decision-makers*; second, that what they need to be able to decide and justify are *the cases that actually arise within that particular society*, and then only in a manner consistent with *past cases that actually arose in that society*; and third, that the decisions need to be justified in terms that have to make sense to *the members of the society at the time*.

If the three parameters are set this way, the need for systematicity mainly arises in the context of public administration or public agencies, and then only in a less demanding form. As a result of keeping both the justifications' scope of application and the circle of their addressees narrowly concrete, less systematization is required, and thicker

conceptual resources are available: a justification applicable to as many cases as necessary but as few as possible can remain thicker and more particularized than one that is applicable to any conceivable case, and even in a highly pluralistic society, there is still far more shared confidence in concepts that can act as a common basis for reason-giving than there would be between beings who shared nothing but the faculty of practical reason.

By Dworkin's lights, the prime example of a public decision-making body dedicated to resolving fundamental conflicts of values by discursively setting them out and debating them 'as issues of principle' (Dworkin 1985, 70) is the U.S. Supreme Court. And here, the systematizing pursuit of theoretical virtues has a place, and one can see what would have led Dworkin to advocate 'justice for hedgehogs', as the title of one his books has it (2011).

But even the U.S. Supreme Court is not subjected to the need to systematize in a maximally strong form. It only addresses cases that actually arise within U.S. jurisdiction, raise issues of constitutional interpretation, and pass various procedural hurdles; it aims to justify its decisions in terms that cover only as many cases as necessary; and since its discursive justifications are addressed to the citizens of a concrete political community, even the Supreme Court can rely on some measure of contingent agreement among them, even if it is only what Cass Sunstein calls *incompletely theorized agreement*—agreement on the value of some grand abstraction without agreement on how exactly it should be specified or realized, for example, or agreement on what the upshot or outcome should be without agreement on the reasons why it should be the upshot or outcome.³¹ Well-functioning legal systems in pluralistic societies tend to exploit such incompletely theorized agreements among citizens by exercising *judicial minimalism*, preferring to articulate their decisions in terms that are, in Sunstein's terminology, 'shallow rather than deep, and narrow rather than wide' (1996, xii): terms that skirt around the deeper and more contested theoretical issues and decide the case at hand in a way that predetermines or constrains as few other decisions as possible.³² As Sunstein shows in *One Case at a Time: Judicial Minimalism on the Supreme Court* (2001), there are many merits to this

³¹ On incompletely theorized agreements, see Sunstein (1996, xi–xii, 35–60).

³² On judicial minimalism and the contrasts between shallowness and depth on the one hand and narrowness and width on the other, see Sunstein (1996, xii, 44–5, 114–5; 2001).

judicial disposition to keep rulings shallow and narrow instead of digging down to first principles with ramifications for a broad range of cases.

While public decision-making bodies like the Supreme Court do need to resolve fundamental conflicts between value concepts and justify their decisions in terms of explicitly stateable and consistent principles, therefore, that need does not drive them all the way to fully systematized conceptual edifices by which they can adjudicate everything from first principles. They only need to resolve certain conceptual tensions rather than all of them; they only need to resolve them on a case-by-case basis rather than once and for all; they do not need to fall back on universally intelligible concepts, but are free to draw on the local and thick conceptual resources that command allegiance in that jurisdiction at the time; and they need to systematize them only a little, in a way that falls far short of yielding a tidy theory.

Yet this needs-based approach to systematization also goes some way towards accommodating the Dworkinian intuition that the values of liberty and equality need to be reconciled. *Some* people, some of the time, really do *need* to resolve conflicts of liberty and equality in principled and moderately systematic fashion. When an instance of such a conflict comes before a court, the judge must look for a way to resolve the tension, to systematize the concepts and considerations involved, and to discursively justify the decision reached. Hence the longstanding allure of picturing the law as a ‘seamless web’, as one nineteenth-century legal historian famously put it.³³ Even on the needs-based approach, therefore, we are led to concede that when a judge faces a conflict between liberty and equality, the Dworkinian aspiration to resolve the conflict by coming as close as possible to reconciling the two concepts is grounded in a real need.

At the same time, resolving a tension between two concepts in a particular case is not the same as resolving it in principle, by *immunizing* the two concepts against conflict. And even when the tension is resolved in the particular case, it is not necessarily resolved *without loss*, nor in a way that prefigures resolutions of other instances of the conflict. The judge’s decision might form a precedent, but as Sunstein stresses, its ramifications should

³³ See Maitland (1898, 13), who primarily applied the image to history, though in a way that implied a similar interconnectedness in the law. Dworkin himself applies the image to the law, though he claims not that the law *is* a seamless web, but that the ideal judge should treat it *as if* it were; see Dworkin (1977, 115–16).

be kept as minimal as possible.

Even if our concerns sometimes give rise to the need to systematize our concepts to some degree, moreover, this need must still be balanced against competing needs arising from countervailing concerns—it will not simply override or silence them. The systematization of thought may therefore come at a price even when it serves some of our concerns, and it certainly comes at the expense of the satisfaction of other concerns when it is pursued beyond need. As Williams remarks of the concepts in terms of which we articulate our moral sentiments, ‘to demand that they be schooled by the requirement of system is to alter our moral perception of the world, not just to make it in some incontestable sense more rational’ (2006h, 162).

Moreover, an outlook that is maximally regimented by the requirements of systematicity and realizes a slew of theoretical virtues is not just a particular outlook, but a particularly trimmed and pared-down one. There is an inevitable trade-off between the rationalization and systematization of thought and the cultivation of its density, texture, and richness. Some degree of systematization may be a practical requirement on public agencies in pluralistic and liberal democratic societies; but extending these demands on administrative forms of reasoning to personal deliberation threatens to flatten and impoverish personal deliberation and experience. There is a danger here of what C. Thi Nguyen has called *value capture*, where thinner, simpler, more widely accessible, and more aggregable conceptions of value end up displacing thicker, richer, subtler, and more local conceptions of value.³⁴ The simplified value concepts may be seductively clear and better tailored to the needs of public administration; but just because of this, they are likely to be less well tailored to the needs of individuals and their private deliberation. And if the thinned-out conceptions tailored to public discourse encroach sufficiently on personal deliberation, individuals risk losing touch with the more complex conceptions of value that sustain their ethical lives, and whose importance in those lives motivated their incorporation into public discourse in the first place.

The worry here is not just that value capture would involve an epistemic loss in the diversity and richness of thought; it is also an ethical worry: that these thinned-out conceptions of value cannot provide enough substance to sustain a worthwhile kind of

³⁴ See Nguyen (2020, 200–203).

life.³⁵ As Williams articulates the point, ‘it is precisely the use of “thick” ethical concepts ... that contributes to a more substantive type of personal ethical experience’ (2005h, 48–9). The ‘intuitive condition’ in which a motley of unsystematized concepts remain in various respects vague, indeterminate, or practically incongruent, he writes,

is not only a state which private understanding *can* live with, but a state which it must have as part of its life, if that life is going to have any density or conviction and succeed in being that worthwhile kind of life which human beings lack unless they feel more than they can say, and grasp more than they can explain. (Williams 1981a, 82)

If systematization comes at a cost to the conceptual breadth, diversity, and richness that gives experience its substance, we should be hesitant to subject all our thought to the systematizing pursuit of theoretically virtues. There is a multi-layered density that is made possible by bringing the full kaleidoscopic jumble of our concepts to bear on our experience. It facilitates a form of comprehensiveness—even, as Nietzsche observes, a form of objectivity, because there is a form of objectivity that does precisely not consist in throwing off whatever is distinctive and parochial about one’s perspective until one anaemically inhabits a view from nowhere, but consists, rather, in taking up and bringing to bear as many different perspectives on a matter as possible: ‘*the more feelings* we allow to come to expression on a matter, *the more eyes*, different eyes, we can use to view the same matter, the more complete will our “concept” of this matter, our “objectivity” be’ (1998, III, §12).

Yet, as we have seen, simply rejecting the demand for systematization is also no longer an option. The bind we are in is that we are subject to conceptual needs that pull in different directions: we need public thought to take one form, personal thought to take another, and the encroachment of either style of thought on the other entrains a kind of loss.³⁶

³⁵ See also Cueni and Queloz (2021) and Harcourt (Manuscript).

³⁶ Hegelian theorists of ‘the virtuous republic’ offer examples of attempts to model public administrative reasoning on personal virtue ethics, while direct utilitarianism exemplifies the opposite tendency to model personal on administrative reasoning (Williams 2005h, 50). Harcourt (Manuscript), drawing on Adorno, illuminatingly explores some further ways in which administrative reasoning encroaches on personal life as its metrics and standards are internalized by private individuals; and so, coming at it from a different direction, does Nguyen (2020).

Once this is recognized, the rationalistic conception of rationality on which rationality inherently requires systematization can itself be recognized as such an encroachment. The conception is encouraged by importing into personal deliberation a demand that has its proper place in public deliberation: the rise of ethics committees, commissions, and panels is not the faithful institutionalization of an antecedent ideal of rationality, but rather what fosters this ideal in the first place. Hence, as Williams notes in an echo of Nietzsche, it is a reversal of cause and effect to view the increasing delegation of decisions of intimate significance to hospital ethics committees and comparable institutional bodies as ‘the fulfilment of an Enlightenment dream, the regulation of ultimate questions by the institutional embodiment of systematic ethical reason’ (2005h, 46):³⁷ it is the conception of practical reason that is being shaped by liberal democratic requirements on public administration, not the other way round.

To recapitulate: modernity combines the growth of public decision-making with the concern that public decision-making be verifiably based on good reasons. This engenders the need for decisions to be discursively justified in principled and consistent terms that make sense to those to whom the decision-makers are answerable. But under conditions of pluralism, those addressees differ widely in their thicker, more substantive concepts. As a result, there is a need for public reasoning to take a more systematic and discursively justifiable form than personal reasoning, and to be articulated in terms that are thinner and more formal or procedural. This serves the concern to hold public decision-makers accountable. But to the extent that this ideal of public reasoning also influences people’s conception of rationality in personal reasoning, it risks entraining a loss of substance in personal thought. The systematization of practical reasoning that has a rightful place in public contexts threatens, when needlessly generalized beyond its proper remit, the conceptual breadth, variety, and richness that is an essential component of the breadth, variety, and richness of experience itself.

On the needs-based approach, the way out of this distinctively modern bind, where we face diverging demands in the personal and public sphere, lies in recognizing that we have different conceptual needs in different domains. The most helpful conceptual apparatus is likely to be a *patchwork* of thicker and thinner concepts, of more socio-

³⁷ As Williams acknowledges, this is a prime example of what Nietzsche described as one of the four great errors, the ‘*error of confusing cause and effect*’ (2005c, Errors, §1).

culturally distinctive and more widely shared concepts.

Given such a conceptual patchwork view, it would be a mistake to think that we first need to choose, say, between the thick concepts of virtues and vices favoured by ancient ethics, the thin deontological concepts favoured by Kantianism, and the equally thin consequentialist concepts favoured by utilitarianism; and that once we have chosen one set of concepts, we should deploy that one set, and only that one set, across the board, to govern our personal lives as well as our public and institutional discourse.

There are different practical pressures acting on the way we think in different contexts, and the concepts that serve us best in our personal lives are not necessarily the same as those that serve us best in positions of public decision-making. We should not aim to find a single, all-purpose set of concepts that we can apply uniformly across different areas of life. And even where we have reason to systematize and move towards concepts at the thinner end of the spectrum, we should do so only to the extent required by the degree of plurality that actually obtains among those to whom the decision-making is answerable. We should aim to think in terms that are as thin and unified as they must be when they must be, but as thick and varied as they can be when they can be.

Conclusion

By inviting us to discern reasons for concept use in our conceptual needs, the approach presented in this book gives us a new way of looking at concepts. In coming to perceive the dense array of needs our concepts are enmeshed in, we come to see that we have more reasons than we knew to conceptualize the world on some lines rather than on others. Considering what concepts we now need allows us to identify the proper place of efforts to tidy up thought; it helps us to adjudicate between competing conceptions of concepts like *voluntariness* or *liberty*; and, more generally, it puts us in a position to decide which concepts to revise, retain, or reject.

What is more, it does all this while reconciling us to the contingency of our concepts: where our thick normative concepts are concerned, certainly, it makes the urge to wring ourselves and our local peculiarities out of our concepts appear misguided. The parochial features of our situation are not distortions to be overcome, but what our concepts are primarily answerable to. Our concepts may be causally contingent, but once it is recognized that the standards they must meet are similarly subject to causal contingency, the sense of the rational contingency of our concepts can be dispelled and replaced by a sense of their necessity given our needs.

The result is a conceptual ethics that is as much about extraconceptual reality as it is about concepts. It does not fall into the trap, which has been looming since the linguistic turn, of thinking that every problem is a conceptual problem that can be solved by tweaking the concepts through which we perceive it. The needs-based approach gives us the means to recognize that solving a problem merely at the conceptual level is often not to solve it at all, but to aggravate the problem by rendering us conceptually blind to it and

its costs. Which concepts we use may often be a crucial aspect of a situation, but it is only one aspect of it, and one that needs to be responsive to the other aspects of the situation. Our concepts are answerable to something outside themselves. We cannot choose them at will.

Nor is the idea that our otherwise rationally undetermined choices of concepts are constrained at the edges by ‘merely pragmatic’ considerations arising from limitations of human nature or contingent features of the world we inhabit. When philosophers have paid any heed to ‘pragmatic’ considerations at all, they have tended to think of them as *constraints* on concept choice: constraints that impose outer limits on what our conceptual scheme could possibly come to look like, but otherwise leave it completely undetermined.¹ Within the framework I have developed, by contrast, the kinds of contingent instrumental considerations that normally get brushed aside and consigned to the leftover category of the ‘merely pragmatic’ are no longer an afterthought, but figure front and center, differentiating into a rich set of intellectual resources that include, alongside limitations by human nature and constraints by physical laws, the local concerns we identify with, the distinctive elaborations of our capacities and circumstances, and the conceptual needs that result from their combination in a particular context.

Conceptual needs do not merely constrain our choices of concepts, moreover, but can positively inform and guide them. As long as they are merely thought of as constraints, considerations of a practical sort can only ever explain why our concepts never seem to stray beyond certain boundaries. But constraints could not give us reasons to revise a concept we already have, or introduce a concept we lack. Thinking of practical considerations as constraints therefore itself constrains our thinking, because it leaves out the most interesting ways in which practical pressures can rationally determine the proper shape of our concepts. The present account fills this blind spot. Practical considerations figure in it not just negatively, as constraints *ruling out* certain concepts, but positively, as reasons *for* concept use.

Becoming sensitive to these reasons itself requires a certain concept, namely the concept of a second-order reason. That concept is indispensable to recognizing the need for, and the intelligibility of, reasons for concept use. The benefit of becoming capable of

¹ See, e.g., Hirsch (1993, 115–116) and Forster (2017, 271; 2004, 67–81).

ascending to the metaconceptual level is that it enables us to raise, from within our practices of reason-giving, the question of whether our reason-giving distinctions can themselves be grounded in reasons. Demanding reasons for reasons transforms the reason relations that govern our reasoning into *objets* of that reasoning: we treat the recognized patterns of correct reasoning that enable us to reason in the first place as themselves standing in need of reasons. This makes it possible to assess and, where necessary, revise our conceptual architecture from within. The concept of a second-order reason thus itself turns out to be one of the more needful concepts in our repertoire. The capacity to conceive of second-order reasons makes it possible to critically examine which first-order reasons we should be responsive to.

In making the case that it is in our conceptual needs that we should look for second-order reasons, I have, in effect, articulated *third-order* reasons: reasons to *count* conceptual needs as second-order reasons to heed certain first-order reasons. To this end, I have also advanced third-order reasons *against* counting the alternative considerations cited by foundationalism, ironism, and holism as second-order reasons: the first, when generalized to cover thick normative concepts, has become incredible; the second results in indiscriminate disengagement from our concepts; and the third results in undiscriminating acceptance of them. To escape this trilemma and find more critical leverage by which to discriminate between concepts that merit confidence and concepts that do not, I have argued that the picture of our conceptual apparatus as something harmonious, largely tensionless, and inherently static must be replaced with a kaleidoscopic picture on which our conceptual apparatus is tension-ridden and dynamic; and that the critical leverage of local needs must be harnessed by recognizing that the contingency of our concepts extends also to the standards that these concepts must meet.

In the case of the most direct rival to the needs-based approach, the tidy-minded pursuit of conceptual authority through the realization of theoretical virtues, the verdict is more nuanced: there is indeed a place for it, but theoretical virtues are not best understood as being themselves what confers authority on a concept—it is rather that our conceptual needs sometimes render theoretically virtuous concepts needful; but equally, our conceptual needs sometimes call for different kinds of concepts, including concepts that are appropriately vague, superficial, or conflictual.

Ultimately, of course, there is only so much that living by the right concepts can, just

by itself, achieve. Once it has been settled *that* a concept is the one we have most reason to use, the question remains of *how* it should be used: when it as opposed to other concepts one also has good reasons to be an engaged user of or should figure in one's deliberations; how exactly the general concept is to be concretely applied to a particular situation; and how the still somewhat indeterminate and often conflicting demands that the situation, thus conceptualized, makes on us should be further concretized, weighed against each other, and acted on. The network of reason relations encoded by our concepts is one thing; our actual practice of reasoning is another. Here, we cannot simply rely on the virtues of concepts, whether theoretical or practical, but must rely on the virtues of *concept-users*—and that means flesh-and-blood human beings, who, for all the good reasons adverted to by the carefully selected mindware at their disposal, retain a mind of their own.

Downstream of conceptual ethics as the reflection on which concepts we have most reason to use, then, lies conceptual ethics as the reflection on how best to use those concepts, which is itself but a tributary to the deeper waters of ethics and politics. The task of conceptual ethics does not end with the specification of the right concepts for us, just as the task of building a well-functioning state does not end with the specification of the right laws and regulations for that state.

In the end, concepts cannot—and need not—do all the work. But they do a great deal of it, and rather more than they are given credit for. By identifying the concepts we need, we can ensure that they do the right kind of work, and do it well.

Bibliography

- Adamson, Peter. 2022. *Don't Think for Yourself: Authority and Belief in Medieval Philosophy*. Notre Dame: University of Notre Dame Press.
- Adorno, Theodor W., Max Horkheimer, and Eugen Kogon. 1989. 'Die verwaltete Welt oder: Die Krisis des Individuum'. In *Max Horkheimer, Gesammelte Schriften Bd. 13: Nachgelassene Schriften 1949–72*, 121–142. Frankfurt am Main: Fischer.
- Aeschylus. 1966. *The Oresteia*. Translated by Robert Fagles. London: Penguin.
- Alanen, Lilli. 2002. 'Descartes on the Will and the Power to Do Otherwise'. In *Emotions and Choice from Boethius to Descartes*. Edited by Henrik Lagerlund and Mikko Yrjönsuuri, 279–298. Dordrecht: Springer.
- Alshanetsky, Eli. 2019. *Articulating a Thought*. Oxford: Oxford University Press.
- Altham, J. E. J. 1995. 'Reflection and Confidence'. In *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*. Edited by J. E. J. Altham and Ross Harrison, 156–69. Cambridge: Cambridge University Press.
- Alvarez, Maria. 2010. *Kinds of Reasons: An Essay in the Philosophy of Action*. Oxford: Oxford University Press.
- Alvarez, Maria, and John Hyman. 2019. 'Philosophy of Action 1945–2015'. In *The Cambridge History of Philosophy, 1945–2015*. Edited by Kelly Becker and Iain D. Thomson, 103–114. Cambridge: Cambridge University Press.
- Anderson, Christopher J., André Blais, Shaun Bowler, Todd Donovan, and Ola Listhaug, eds. 2005. *Losers' Consent: Elections and Democratic Legitimacy*. Oxford: Oxford University Press.
- Anderson, Elizabeth. 1995. 'Knowledge, Human Interests, and Objectivity in Feminist Epistemology'. *Philosophical Topics* 23 (2): 27–58.
- . 2001. 'Unstrapping the Straitjacket of 'Preference': A Comment on Amartya Sen's Contributions to Philosophy and Economics'. *Economics and Philosophy* 17 (1): 21–38.
- . 2004. 'Uses of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce'. *Hypatia* 19 (1): 1–24.
- Anscombe, G. E. M. 2005. 'Action, Intention and "Double Effect"'. In *Faith in a Hard Ground: Essays on Religion, Philosophy and Ethics by G.E.M. Anscombe*. Edited by Mary Geach and Luke Gormally, 207–26. Exeter: Imprint Academic.
- . 2008a. 'On Being in Good Faith'. In *Faith in a Hard Ground: Essays on Religion, Philosophy and Ethics by G.E.M. Anscombe*. Edited by Mary Geach and Luke Gormally, 101–112. Exeter: Imprint Academic.
- . 2008b. 'Sin: the McGivney lectures'. In *Faith in a Hard Ground: Essays on Religion, Philosophy and Ethics by G.E.M. Anscombe*. Edited by Mary Geach and Luke Gormally, 117–156. Exeter: Imprint Academic.
- Anscombe, Gertrude Elizabeth Margaret. 1957. *Intention*. Oxford: Blackwell.
- Appiah, Kwame Anthony. 2010. *The Honor Code: How Moral Revolutions Happen*. New York: W. W. Norton.
- Arendt, Hannah. 1956. 'Authority in the Twentieth Century'. *The Review of Politics* 18 (4): 403–417

- . 1958. ‘What Was Authority?’. *NOMOS: American Society for Political and Legal Philosophy* 1: 81–112.
- . 1968. ‘Truth and Politics’. In *Between Past and Future: Eight Exercises in Political Thought*, 227–64. New York: Viking.
- Ariew, Andre, Robert Cummins, and Mark Perlman. 2002. *Functions: New Essays in the Philosophy of Psychology and Biology*. Oxford: Oxford University Press.
- Asgeirsson, Hrafn. 2015. ‘On the Instrumental Value of Vagueness in the Law’. *Ethics* 125 (2): 425–448.
- . 2020. *The Nature and Value of Vagueness in the Law*. Oxford: Hart.
- Austin, John. 1885. *Lectures on Jurisprudence: Or, The Philosophy of Positive Law*. Edited by Robert Campbell. 5th ed. London: John Murray.
- Austin, John Langshaw. 1961. ‘A Plea for Excuses’. In *Philosophical Papers*. Edited by J. O. Urmson and G. J. Warnock, 123–52. Oxford: Clarendon Press.
- . 1962. *Sense and Sensibilia*. Edited by G. J. Warnock. Oxford: Oxford University Press.
- Babiotti, Paolo. 2020. ‘Compression: Nietzsche, Williams, and the Problem of Style’. *European Journal of Philosophy*.
- Bader, Ralf M. 2015. ‘Kantian Axiology and the Dualism of Practical Reason’. In *The Oxford Handbook of Value Theory*. Edited by Iwao Hirose and Jonas Olson, 175–202. Oxford: Oxford University Press.
- Baillie, James. 2000. *Hume on Morality*. London: Routledge.
- Balkin, J. L. 1998. *Cultural Software: A Theory of Ideology*. New Haven: Yale University Press.
- Ball, Derek. 2020. ‘Metasemantic Ethics’. *Ratio* 33 (4): 206–219.
- Ball, Terence, James Farr, and Russell L. Hanson, eds. 1989. *Political Innovation and Conceptual Change*. New York: Cambridge University Press.
- Barber, Nicholas. 2010. *The Constitutional State*. Oxford: Oxford University Press.
- Baron, Marcia. 2003. ‘Manipulativeness’. *Proceedings and Addresses of the American Philosophical Association* 77 (2): 37–54.
- Baz, Avner. 2017. *The Crisis of Method in Contemporary Analytic Philosophy*. Oxford: Oxford University Press.
- Beebe, James R. 2012. ‘Social Functions of Knowledge Attributions’. In *Knowledge Ascriptions*. Edited by Jessica Brown and Mikkel Gerken, 220–242. Oxford: Oxford University Press.
- Beisbart, Claus, and Georg Brun. manuscript. ‘How to Defend Reflective Equilibrium’.
- Benedict, Ruth. 1934. *Patterns of Culture*. New York: Penguin.
- Bennett, Jonathan. 1976. *Linguistic Behaviour*. Cambridge: Cambridge University Press.
- . 2008. ‘Accountability (II)’. In *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson’s “Freedom and Resentment”*. Edited by Michael McKenna and Paul Russell, 47–68. Farnham: Ashgate.
- Bentham, Jeremy. 1843. *The Works of Jeremy Bentham, Vol. 2*. Edinburgh: William Tait.
- Berlin, Isaiah. 2002a. ‘The Birth of Greek Individualism: A Turning-Point in the History of Political Thought’. In *Liberty*. Edited by Henry Hardy, 287–321. Oxford: Oxford University Press.
- . 2002b. ‘Two Concepts of Liberty’. In *Liberty*. Edited by Henry Hardy, 166–217. Oxford: Oxford University Press.
- . 2013a. ‘The Decline of Utopian Ideas in the West’. In *The Crooked Timber of Humanity: Chapters in the History of Ideas*. Edited by Henry Hardy, 21–50. Princeton: Princeton University Press.

- . 2013b. ‘European Unity and Its Vicissitudes’. In *The Crooked Timber of Humanity: Chapters in the History of Ideas*. Edited by Henry Hardy, 186–218. Princeton: Princeton University Press.
- . 2013c. ‘My Intellectual Path’. In *The Power of Ideas*. Edited by Henry Hardy. 2nd ed, 1–28. Princeton: Princeton University Press.
- . 2013d. ‘The Pursuit of the Ideal’. In *The Crooked Timber of Humanity: Chapters in the History of Ideas*. Edited by Henry Hardy, 1–20. Princeton: Princeton University Press.
- . 2014a. ‘*Helvétius*’. In *Freedom and Its Betrayal: Six Enemies of Human Liberty*. Edited by Henry Hardy, 11–27. Princeton: Princeton University Press.
- . 2014b. ‘Politics as a Descriptive Science’. In *Political Ideas in the Romantic Age: Their Rise and Influence on Modern Thought*. Edited by Henry Hardy, 21–111. Princeton: Princeton University Press.
- . 2015. *Affirming: Letters 1975–1997*. Edited by Henry Hardy and Mark Pottle. London: Chatto and Windus.
- Berlin, Isaiah, and Bernard Williams. 1994. ‘Pluralism and Liberalism: A Reply’. *Political Studies* 42 (2): 306–309.
- Bermúdez, José Luis. 2021. *Frame It Again: New Tools for Rational Decision-Making*. New York: Cambridge University Press.
- Bernays, Edward. 1969. *The Engineering of Consent*. Norman: University of Oklahoma Press.
- Berson, Josh. 2021. *The Human Scaffold: How Not to Design Your Way Out of a Climate Crisis*. Oakland, CA: University of California Press.
- Besson, Corinne. 2018. ‘Norms, Reasons, and Reasoning: A Guide Through Lewis Carroll’s Regress Argument’. In *The Oxford Handbook of Reasons and Normativity*. Edited by Daniel Star, 504–528. New York: Oxford University Press.
- . forthcoming. *Logic, Reasoning and Lewis Carroll*. Oxford: Oxford University Press.
- Bierson, Marshall, and John Schwenkler. 2022. ‘What is the Bearing of Thinking on Doing?’. In *The Anscombean Mind*. Edited by Adrian Haddock and Rachael Wiseman, 312–32. Abingdon: Routledge.
- Binmore, Ken. 2005. *Natural Justice*. New York: Oxford University Press.
- . 2007. *Playing for Real: A Text on Game Theory*. Oxford: Oxford University Press.
- Blackburn, Simon. 1993. *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- . 1998. *Ruling Passions*. Oxford: Oxford University Press.
- . 2005. ‘Success Semantics’. In *Ramsey’s Legacy*. Edited by Hallvard Lillehammer and D. H. Mellor, 22–36. Oxford: Oxford University Press.
- . 2013a. ‘Pragmatism in Philosophy: The Hidden Alternative’. *Philosophic Exchange* 41 (1): 2–13.
- . 2013b. ‘Pragmatism: All or Some?’. In *Expressivism, Pragmatism and Representationalism*. Edited by Huw Price, 67–84. Cambridge: Cambridge University Press.
- . 2017. ‘Pragmatism: All or Some or All and Some?’. In *The Practical Turn: Pragmatism in Britain in the Long Twentieth Century*. Edited by Cheryl Misak and Huw Price, 61–74. Oxford: Oxford University Press.
- . 2019. ‘Lonely in Littlemore: Confidence in Ethics and the Limits of Philosophy’. In *Ethics Beyond the Limits: New Essays on Bernard Williams’ Ethics and the Limits of Philosophy*. Edited by Sophie Grace Chappell and Marcel van Ackeren, 27–36. London: Routledge.

- Blome-Tillmann, Michael. 2009. 'Contextualism, Subject-Sensitive Invariantism, and the Interaction of 'Knowledge'-Ascriptions with Modal and Temporal Operators*'. *Philosophy and Phenomenological Research* 79 (2): 315–31.
- Bloor, David. 1983. *Wittgenstein: A Social Theory of Knowledge*. New York: Columbia University Press.
- . 1992. 'Left and Right Wittgensteinians'. In *Science as Practice and Culture*. Edited by Andrew Pickering, 266–82. Chicago: University of Chicago Press.
- . 1997. *Wittgenstein, Rules and Institutions*. London: Routledge.
- . 2000. 'Wittgenstein as a Conservative Thinker'. In *The Sociology of Philosophical Knowledge*. Edited by Martin Kusch, 1–14. Dordrecht: Springer Netherlands.
- . 2004. 'Ludwig Wittgenstein and Edmund Burke'. In *Essays on Wittgenstein and Austrian Philosophy: In Honour of J.C. Nyiri*. Edited by Tamás Demeter, 109–34. Amsterdam: Rodopi.
- Blumenthal, Geoffrey, and James Ladyman. 2017. 'The Development of Problems Within the Phlogiston Theories, 1766–1791'. *Foundations of Chemistry* 19 (3): 241–280.
- Boghossian, Paul. 2003. 'Blind Reasoning'. *Aristotelian Society Supplementary Volume* 77 (1): 225–248.
- Bowles, Samuel, and Hebert Gintis. 2011. *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton: Princeton University Press.
- Brandom, Robert. 1994. *Making It Explicit. Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.
- . 2000. *Articulating Reasons*. Cambridge, MA: Harvard University Press.
- . 2001. 'Reason, Expression, and the Philosophic Enterprise'. In *What is Philosophy?* Edited by C. P. Ragland and Sarah Heidt, 74–95. New Haven: Yale University Press.
- . 2002a. *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality*. Cambridge, MA: Harvard University Press.
- . 2002b. 'When Philosophy Paints Its Blue on Gray: Irony and the Pragmatist Enlightenment'. *boundary 29* (2): 1–28.
- . 2004. 'The Pragmatist Enlightenment (and Its Problematic Semantics)'. *European Journal of Philosophy* 12 (1): 1–16.
- . 2008. *Between Saying and Doing*. Oxford: Oxford University Press.
- . 2009. *Reason in Philosophy: Animating Ideas*. Cambridge, MA: Belknap Press of Harvard University Press.
- . 2011. 'Vocabularies of Pragmatism: Synthesizing Naturalism and Historicism'. In *Perspectives on Pragmatism: Classical, Recent, and Contemporary*, 116–157. Cambridge, MA: Harvard University Press.
- . 2013. 'Global Anti-representationalism?'. In *Expressivism, Pragmatism and Representationalism*. Edited by Huw Price, 85–111. Cambridge: Cambridge University Press.
- . 2014. 'A Hegelian Model of Legal Concept Determination: The Normative Fine Structure of the Judges' Chain Novel'. In *Pragmatism, Law, and Language*. Edited by Graham Hubbs and Douglas Lind, 19–39. New York: Taylor and Francis.
- . 2015a. 'Den Abgrund reflektieren: Vernunft, Genealogie und die Hermeneutik des Edelmuts'. *West End: Neue Zeitschrift für Sozialforschung* 1: 3–26.
- . 2015b. *From Empiricism to Expressivism: Brandom Reads Sellars*. Cambridge, MA: Harvard University Press.

- . 2019a. ‘Some Strands of Wittgenstein’s Normative Pragmatism, and Some Strains of his Semantic Nihilism’. *Disputatio* 8 (9): 1–29.
- . 2019b. *A Spirit of Trust: A Reading of Hegel’s Phenomenology*. Cambridge, MA: Harvard University Press.
- . 2021. ‘Achieving the Enlightenment’. In *Pragmatism as Anti-Authoritarianism*. Edited by Eduardo Mendieta, vii–xxvi. Cambridge, MA: Belknap Press.
- Brigandt, Ingo. 2006. *A Theory of Conceptual Advance: Explaining Conceptual Change in Evolutionary, Molecular, and Evolutionary Developmental Biology*. Doctoral Dissertation, University of Pittsburgh, <http://d-scholarship.pitt.edu/8849/>.
- . 2010. ‘The Epistemic Goal of a Concept: Accounting for the Rationality of Semantic Change and Variation’. *Synthese* 177 (1): 19–40.
- . 2011. ‘Natural Kinds and Concepts: A Pragmatist and Methodologically Naturalistic Account’. In *Pragmatism, Science and Naturalism*. Edited by Jonathan Knowles and Henrik Rydenfelt, 171–96. Frankfurt a. M.: Peter Lang.
- . 2012. ‘The Dynamics of Scientific Concepts: The Relevance of Epistemic Aims and Values’. In *Scientific Concepts and Investigative Practice*. Edited by Uljana Feest and Friedrich Steinle, 75–103. Berlin: De Gruyter.
- . 2013. ‘A Critique of David Chalmers’ and Frank Jackson’s Account of Concepts’. *ProtoSociology* 30: 63–88.
- Brigandt, Ingo, and Esther Rosario. 2020. ‘Strategic Conceptual Engineering for Epistemic and Social Aims’. In *Conceptual Engineering and Conceptual Ethics*. Edited by Alexis Burgess, Herman Cappelen and David Plunkett, 100–124. Oxford: Oxford University Press.
- Brown, Thomas. 2012. *Inquiry into the Relation of Cause and Effect*. Cambridge: Cambridge University Press.
- Brun, Georg. 2014. ‘Reflective Equilibrium Without Intuitions?’. *Ethical Theory and Moral Practice* 17 (2): 237–252.
- . 2016. ‘Explication as a Method of Conceptual Re-Engineering’. *Erkenntnis* 81 (6): 1211–1241.
- . 2020. ‘Conceptual Re-engineering: From Explication to Reflective Equilibrium’. *Synthese* 197 (3): 925–54.
- . 2022. ‘Re-Engineering Contested Concepts. A Reflective-Equilibrium Approach’. *Synthese* 200 (2): 1–29.
- Buchler, Justus. 1955. *Nature and Judgment*. New York: Columbia University Press.
- Burge, Tyler. 1979. ‘Individualism and the Mental’. *Midwest Studies in Philosophy* 4 (1): 73–122.
- Burgess, Alexis, and David Plunkett. 2013a. ‘Conceptual Ethics I’. *Philosophy Compass* 8 (12): 1091–1101.
- . 2013b. ‘Conceptual Ethics II’. *Philosophy Compass* 8 (12): 1102–1110.
- Buss, Sarah. 2005. ‘Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints’. *Ethics* 115 (2): 195–235.
- Byron, George Gordon, Lord. 2015. *Manfred: An Edition of Byron’s Manuscripts and a Collection of Essays*. Edited by Peter Cochran. Cambridge: Cambridge Scholars Publishing.
- Campbell, John Keim, Michael O’Rourke, and Matthew H. Slater, eds. 2011. *Carving Nature at Its Joints: Natural Kinds in Metaphysics and Science*. Cambridge, MA: MIT Press.

- Candlish, Stewart, and Nic Damnjanovic. 2013. 'Reasons, Actions, and the Will: The Fall and Rise of Causalism'. In *The Oxford Handbook of the History of Analytic Philosophy*. Edited by Michael Beaney, 689–708. Oxford: Oxford University Press.
- Cappelen, Herman. 2013. 'Nonsense and Illusions of Thought'. *Philosophical Perspectives* 27 (1): 22–50.
- . 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.
- . 2020. 'Conceptual Engineering: The Master Argument'. In *Conceptual Engineering and Conceptual Ethics*. Edited by Alexis Burgess, Herman Cappelen and David Plunkett, 132–151. Oxford: Oxford University Press.
- Cappelen, Herman, and David Plunkett. 2020. 'Introduction: A Guided Tour of Conceptual Engineering and Conceptual Ethics'. In *Conceptual Engineering and Conceptual Ethics*. Edited by Alexis Burgess, Herman Cappelen and David Plunkett, 1–26. Oxford: Oxford University Press.
- Carey, Susan. 1985. *Conceptual Change in Childhood*. Cambridge, MA: MIT Press.
- Carnap, Rudolf. 1947. *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.
- . 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.
- . 1952. 'Meaning Postulates'. *Philosophical Studies* 3 (5): 65–73.
- . 1962. *Logical Foundations of Probability*. 2nd ed. Chicago: The University of Chicago Press.
- Carroll, Lewis. 1895. 'What the Tortoise Said to Achilles'. *Mind* IV (14): 278–280.
- Carruthers, Peter. 1987. 'Conceptual Pragmatism'. *Synthese* 73 (2): 205–224.
- Carus, André. 2007. *Carnap and Twentieth-Century Thought: Explication as Enlightenment*. Cambridge: Cambridge University Press.
- Carus, André W. 2017. 'Carnapian Rationality'. *Synthese* 194 (1): 163–184.
- Celikates, Robin. 2015. 'Against Manichaeism: The Politics of Forms of Life and the Possibilities of Critique'. *Raisons politiques* 57 (1): 81–96.
- . 2018. *Critique as Social Practice: Critical Theory and Social Self-Understanding*. London: Rowman and Littlefield.
- Chadha-Sridhar, Ira. 2021. 'The Value of Vagueness: A Feminist Analysis'. *Canadian Journal of Law and Jurisprudence* 34 (1): 59–84.
- Chalmers, David J. 2002. 'On Sense and Intension'. *Philosophical Perspectives* 16: 135–82.
- Chang, Hasok. 2004. *Inventing Temperature: Measurement and Scientific Progress*. Oxford: Oxford University Press.
- Chang, Ruth, ed. 1997. *Incommensurability, Incomparability, and Practical Reason*. Cambridge, MA: Harvard University Press.
- . 2002. *Making Comparisons Count*. London: Routledge.
- . 2009. 'Voluntarist Reasons and the Sources of Normativity'. In *Reasons for Action*. Edited by David Sobel and Steven Wall, 243–271. Cambridge: Cambridge University Press.
- . 2015. 'Value Incomparability and Incommensurability'. In *The Oxford Handbook of Value Theory*. Edited by Iwao Hirose and Jonas Olson, 205–224. Oxford: Oxford University Press.
- . 2016. 'Comparativism: The Grounds of Rational Choice'. In *Weighing Reasons*. Edited by Errol Lord and Barry Maguire, 213–240. New York: Oxford University Press.

- Chappell, Sophie Grace, ed. 2015. *Intuition, Theory, and Anti-Theory in Ethics*. Oxford: Oxford University Press.
- Chappell, Timothy. 2009. 'Ethics Beyond Moral Theory'. *Philosophical Investigations* 32 (3): 206–243.
- Chesebro, James W. 1985. 'Definition as a Rhetorical Strategy'. *Pennsylvania State Communication Annual* 45 (1): 5–15.
- Chihara, Charles. 1979. 'The Semantic Paradoxes: A Diagnostic Investigation'. *The Philosophical Review* 88 (4): 590–618.
- Christiano, Tom. 2020. 'Authority'. In *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Summer 2020 edition ed.
- Churchland, Patricia. 1986. *Neurophilosophy*. Cambridge, MA: MIT Press.
- Clark, Andy. 2013. *Mindware: An Introduction to the Philosophy of Cognitive Science*. New York: Oxford University Press.
- Clarke-Doane, Justin. 2020. *Morality and Mathematics*. Oxford: Oxford University Press.
- Cohen, Jean L. 2012. *Globalization and Sovereignty: Rethinking Legitimacy, Legitimacy, and Constitutionalism*. Cambridge: Cambridge University Press.
- Cohen, Stewart. 2016. 'Theorizing about the Epistemic'. *Inquiry* 59 (7–8): 839–57.
- Cohon, Rachel. 2008. *Hume's Morality: Feeling and Fabrication*. New York: Oxford University Press.
- Congdon, Matthew. 2023. *Moral Articulation: On the Development of New Moral Concepts*. New York: Oxford University Press.
- Coons, Christian, and Michael Weber. 2014. *Manipulation: Theory and Practice*. New York: Oxford University Press.
- Craig, Edward. 1986. 'The Practical Explication of Knowledge'. *Proceedings of the Aristotelian Society* 87: 211–226.
- . 1990. *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Oxford: Clarendon Press.
- . 1993. *Was wir wissen können: Pragmatische Untersuchungen zum Wissensbegriff. Wittgenstein-Vorlesungen der Universität Bayreuth*. Edited by Wilhelm Vossenkuhl. Frankfurt am Main: Suhrkamp.
- . 2000. 'Response to Lehrer'. *Philosophy and Phenomenological Research* 60 (3): 655–65.
- . 2007. 'Genealogies and the State of Nature'. In *Bernard Williams*. Edited by Alan Thomas, 181–200. Cambridge: Cambridge University Press.
- Crisp, Roger. 2017. 'Moral Luck and Equality of Opportunity'. *Aristotelian Society Supplementary Volume* 91 (1): 1–20.
- Cueni, Damian. 2020. *Public Law Analogies in International Legal Theory*. Doctoral Thesis, Rechtswissenschaftliches Institut, University of Zurich.
- . manuscript-a. 'Constructing Liberty and Equality: Political, Not Juridical'.
- . manuscript-b. 'The Expressive Function of Basic Rights'.
- . manuscript-c. 'The Legal Architecture of Freedom: Reconstructing the Two-Step Framework'.
- . manuscript-d. 'Revitalizing Rights Infringements'.
- Cueni, Damian, and Matthieu Queloz. 2021. 'Whence the Demand for Ethical Theory?'. *American Philosophical Quarterly* 58 (2): 135–46.
- . 2022. 'Theorizing the Normative Significance of Critical Histories for International Law'. *Journal of the History of International Law*.

- Cullity, Garrett. forthcoming. 'Williams, Berlin, and the Vindication Problem'. In *Making Sense of the Past: Bernard Williams and the History of Philosophy*. Edited by Marcel Van ackeren and Matthieu Queloz. Oxford: Oxford University Press.
- D'Oro, Giuseppina, and Constantine Sandis, eds. 2013. *Reasons and Causes: Causalism and Anti-Causalism in the Philosophy of Action*. New York: Palgrave Macmillan.
- Dancy, Jonathan. 1995. 'In Defense of Thick Concepts'. *Midwest Studies In Philosophy* 20 (1): 263–279.
- . 2004. *Ethics without Principles*. Oxford: Clarendon Press.
- . 2017. 'Moral Particularism'. In *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Winter 2017 ed.
- Daniels, Norman. 1979. 'Wide Reflective Equilibrium and Theory Acceptance in Ethics'. *The Journal of Philosophy* 76 (5): 256–282.
- . 1996. *Justice and Justification: Reflective Equilibrium in Theory and Practice*. Cambridge: Cambridge University Press.
- . 2020. 'Reflective Equilibrium'. In *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Summer 2020 ed.: Metaphysics Research Lab, Stanford University.
- Darden, Lindley. 2006. *Reasoning in Biological Discoveries: Essays on Mechanisms, Interfield Relations, and Anomaly Resolution*. Cambridge: Cambridge: Cambridge University Press.
- Darwall, Stephen. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press.
- . 2010. 'Authority and Reasons: Exclusionary and Second-Personal'. *Ethics* 120 (2): 257–278.
- Darwall, Stephen L. 1977. 'Two Kinds of Respect'. *Ethics* 88 (1): 36–49.
- Daston, Lorraine. 2022. *Rules: A Short History of What We Live By*. Princeton: Princeton University Press.
- Davidson, Donald. 1980. *Essays on Actions and Events*. Oxford: Clarendon Press, Oxford University Press.
- . 1990. 'The Structure and Content of Truth'. *Journal of Philosophy* 87 (6): 279–328.
- . 2001. 'Actions, Reasons, and Causes'. In *Essays on Actions and Events*, 3–20. Oxford: Clarendon Press, Oxford University Press.
- Davies, John. 1996. 'A Fugitive from the Pigeonhole'. *Times Higher Education Supplement* November 1 (1252): 15.
- De Jaeger, Kris. 2003. 'A Game-Theoretic Rationale for Vagueness'. *Linguistics and Philosophy* 26 (5): 637–659.
- De Jaeger, Kris, and Robert van Rooij. 2011. 'Strategic Vagueness, and Appropriate Contexts'. In *Language, Games, and Evolution: Trends in Current Research on Language and Game Theory*. Edited by Anton Benz, Christian Ebert, Gerhard Jäger and Robert van Rooij, 40–59. Berlin: Springer.
- De Mesel, Benjamin. 2021. 'Being and Holding Responsible: Reconciling the Disputants through a Meaning-Based Strawsonian Account'. *Philosophical Studies* 179 (6): 1893–1913.
- Deigh, John. 2008. *Emotions, Values, and the Law*. Oxford: Oxford University Press.
- Delacroix, Sylvie. 2022. *Habitual Ethics?* Oxford: Hart.
- Dennett, Daniel Clement. 1989. *The Intentional Stance*. Cambridge, MA: MIT Press.
- Descartes, René. 1996. *Oeuvres de Descartes*. Edited by Charles Adam and Paul Tannery. 11 vols. Vol. I–XI. Paris: Vrin.
- Dewey, John. 1938. *Logic: The Theory of Inquiry*. New York: Henry Holt.

- Diamond, Cora. 1988. 'Losing Your Concepts'. *Ethics* 98 (2): 255–277.
- . 2018. 'Wittgenstein, Mathematics, and Ethics: Resisting the Attractions of Realism'. In *The Cambridge Companion to Wittgenstein*. Edited by Hans Sluga and David G. Stern. 2nd ed, 209–244. Cambridge: Cambridge University Press.
- . 2021. 'Suspect Notions and the Concept Police'. In *Cora Diamond on Ethics*. Edited by Maria Balaska, 7–30. New York: Palgrave.
- Dorsey, Dale. 2016. *The Limits of Moral Authority*. Oxford: Oxford University Press.
- Dretske, Fred. 1981. *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
- Duff, Anthony. 2004. 'Action, the Act Requirement and Criminal Liability'. In *Agency and Action*. Edited by John Hyman and Helen Steward, 69–103. Cambridge: Cambridge University Press.
- Duff, Anthony, and Andrew von Hirsch. 1997. 'Responsibility, Retribution and the Voluntary: A Response to Williams'. *Cambridge Law Journal* 56 (1): 103–13.
- Dummett, Michael. 1973. *Frege: Philosophy of Language*. New York: Harper and Row.
- Dunaway, Billy. 2020. *Reality and Morality*. Oxford: Oxford University Press.
- Dutilh Novaes, Catarina. 2020a. 'Carnap meets Foucault: Conceptual Engineering and Genealogical Investigations'. *Inquiry*: 1–27.
- . 2020b. 'Carnapian Explication and Ameliorative Analysis: A Systematic Comparison'. *Synthese* 197 (3): 1001–34.
- Dutilh Novaes, Catarina, and Erich Reck. 2017. 'Carnapian Explication, Formalisms as Cognitive Tools, and the Paradox of Adequate Formalization'. *Synthese* 194 (1): 195–215.
- Dworkin, Gerald. 1989. *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.
- Dworkin, Ronald. 1977. *Taking Rights Seriously*. Cambridge, MA: Harvard University Press.
- . 1985. *A Matter of Principle*. Cambridge, MA: Harvard University Press.
- . 1986. *Law's Empire*. Cambridge, MA: Harvard University Press.
- . 2000. *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge, MA: Harvard University Press.
- . 2001a. 'Do Liberal Values Conflict?'. In *The Legacy of Isaiah Berlin*. Edited by Mark Lilla, Ronald Dworkin and Robert Silvers, 73–90. New York: New York Review of Books.
- . 2001b. 'Do Values Conflict: A Hedgehog's Approach'. *Arizona Law Review* 43 (2): 251–260.
- . 2011. *Justice for Hedgehogs*. Cambridge, MA: Harvard University Press.
- Dworkin, Ronald, Bernard Williams, Mark Lilla, Thomas Nagel, Richard Wollheim, Frances Kamm, and Steven Lukes. 2001. 'Pluralism'. In *The Legacy of Isaiah Berlin*. Edited by Mark Lilla, Ronald Dworkin and Robert Silvers, 121–139. New York: New York Review of Books.
- Eberhardt, Jennifer L., Philip Attiba Goff, Valerie J. Purdie, and Paul G. Davies. 2004. 'Seeing Black: Race, Crime, and Visual Processing'. *Journal of Personality and Social Psychology* 87 (6): 876–893.
- Eco, Umberto. 1984. *Semiotics and the Philosophy of Language*. Bloomington: Indiana University Press.
- Edmonds, David. 2020. *The Murder of Professor Schlick: The Rise and Fall of the Vienna Circle*. Princeton: Princeton University Press.

- Eklund, Matti. 2002. ‘Inconsistent Languages’. *Philosophy and Phenomenological Research* 64 (2): 251–275.
- . 2007. ‘Meaning-Constitutivity’. *Inquiry* 50 (6): 559–574.
- . 2017. *Choosing Normative Concepts*. Oxford: Oxford University Press.
- . 2019. ‘Inconsistency and Replacement’. *Inquiry* 62 (4): 387–402.
- . 2021. ‘Conceptual Engineering in Philosophy’. In *The Routledge Handbook of Social and Political Philosophy of Language*. Edited by Rachel Sterken and Justin Khoo. London: Routledge.
- Elgin, Catherine Z. 1983. *With Reference to Reference*. Indianapolis: Hackett.
- . 1996. *Considered Judgment*. Princeton: Princeton University Press.
- . 2017. *True Enough*. Cambridge, MA: MIT Press.
- Eliot, George. 1999. *Daniel Deronda*. Edited by John Rignall. London: Everyman.
- Elliott-Graves, Alkistis, and Michael Weisberg. 2014. ‘Idealization’. *Philosophy Compass* 9 (3): 176–185.
- Emilsson, Anton. manuscript. ‘The Unanswered Question of ‘Freedom and Resentment’.
- Endicott, Timothy. 2000. *Vagueness in Law*. Oxford: Oxford University Press.
- Eschenburg, Theodor. 1976. *Über Autorität*. Frankfurt: Suhrkamp.
- Finnis, John. 1985. ‘On ‘Positivism’ and ‘Legal Rational Authority’’. *Oxford Journal of Legal Studies* 5 (1): 74–90.
- Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Flanagan, Owen. 1991. *Varieties of Moral Personality: Ethics and Psychological Realism*. Cambridge, MA: Harvard University Press.
- Fodor, Jerry A. 2003. *Hume Variations*. Oxford: Clarendon Press.
- . 2004. ‘Having Concepts: A Brief Refutation of the Twentieth Century’. *Mind and Language* 19 (1): 29–47.
- Forrester, Katrina. 2019. *In the Shadow of Justice: Postwar Liberalism and the Remaking of Political Philosophy*. Princeton: Princeton University Press.
- Forster, Michael N. 2004. *Wittgenstein on the Arbitrariness of Grammar*. Princeton: Princeton University Press.
- . 2017. ‘The Autonomy of Grammar’. In *A Companion to Wittgenstein*. Edited by Hans-Johann Glock and John Hyman, 269–277. Oxford: Wiley.
- Frankfurt, Harry G. 1984. ‘Necessity and Desire’. *Philosophy and Phenomenological Research* 45 (1): 1–13.
- . 2004. *The Reasons of Love*. Princeton: Princeton University Press.
- Franklin-Hall, Laura R. 2015. ‘Natural Kinds as Categorical Bottlenecks’. *Philosophical Studies* 172 (4): 925–948.
- Fraser, Rachel Elizabeth. 2018. ‘The Ethics of Metaphor’. *Ethics* 28 (4): 728–755.
- Fredericks, Rachel. 2018. ‘Moral Responsibility for Concepts’. *European Journal of Philosophy* 26 (4): 1381–1397.
- . 2020. ‘Moral Responsibility for Concepts, Continued: Concepts as Abstract Objects’. *European Journal of Philosophy*.
- Fricker, Miranda. 1998. ‘Rational Authority and Social Power: Towards a Truly Social Epistemology’. *Proceedings of the Aristotelian Society* 98 (2): 159–177.
- . 2000. ‘Confidence and Irony’. In *Morality, Reflection, and Ideology*. Edited by Edward Harcourt, 87–112. Oxford: Oxford University Press.

- . 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
- . 2010a. ‘Can There Be Institutional Virtues?’. In *Oxford Studies in Epistemology (Special Theme: Social Epistemology)*, Vol. 3. Edited by Tamar Szabo Gendler and John Hawthorne, 235–252. Oxford: Oxford University Press.
- . 2010b. ‘The Relativism of Blame and Williams’s Relativism of Distance’. *Aristotelian Society Supplementary Volume* 84 (1): 151–77.
- . 2010c. ‘Scepticism and the Genealogy of Knowledge: Situating Epistemology in Time’. In *Social Epistemology*. Edited by A. Haddock, A. Millar and D. Pritchard, 51–68. Oxford: Oxford University Press.
- . 2012. ‘Group Testimony? The Making of A Collective Good Informant’. *Philosophy and Phenomenological Research* 84 (2): 249–276.
- . 2013. ‘Styles of Moral Relativism: A Critical Family Tree’. In *The Oxford Handbook of the History of Ethics*. Edited by Roger Crisp, 793–817. Oxford: Oxford University Press.
- . 2016a. ‘Fault and No-Fault Responsibility for Implicit Prejudice: A Space for Epistemic ‘Agent-Regret’’. In *The Epistemic Life of Groups: Essays in the Epistemology of Collectives*. Edited by Michael S. Brady and Miranda Fricker, 33–50. Oxford: Oxford University Press.
- . 2016b. ‘What’s the Point of Blame? A Paradigm Based Explanation’. *Noûs* 50 (1): 165–183.
- . 2019. ‘Forgiveness: An Ordered Pluralism’. *Australasian Philosophical Review* 3 (1): 241–60.
- . 2020a. ‘Bernard Williams as a Philosopher of Ethical Freedom’. *Canadian Journal of Philosophy* 50 (8): 919–33.
- . 2020b. ‘Institutional Epistemic Vices: The Case of Inferential Inertia’. In *Vice Epistemology*. Edited by Ian James Kidd, Heather Battaly and Quassim Cassam, 89–107. London: Routledge.
- . forthcoming. *Blaming and Forgiving: The Work of Morality*. Oxford: Oxford University Press.
- Friedrich, Carl Joachim. 1958. *Authority*. Cambridge, MA: Harvard University Press.
- . 1971. ‘Authority, Reason, and Discretion’. In *Authority and Social Work*. Edited by Shankar A. Yelaja, 17–34. Toronto: University of Toronto Press.
- . 1972. *Tradition and Authority*. London: Pall Mall.
- Gaitán, Antonio, and Hugo Viciana. 2018. ‘Relativism of Distance - a Step in the Naturalization of Meta-Ethics’. *Ethical Theory and Moral Practice* 21 (2): 311-327.
- Galston, William A. 2002. *Liberal Pluralism: The Implications of Value Pluralism for Political Theory and Practice*. Cambridge: Cambridge University Press.
- . 2005. *The Practice of Liberal Pluralism*. Cambridge: Cambridge University Press.
- Gauker, Christopher. 2011. *Words and Images: An Essay on the Origin of Ideas*. Oxford: Oxford University Press.
- Geertz, Clifford. 1973a. ‘Ideology as a Cultural System’. In *The Interpretation of Cultures: Selected Essays*, 193–233. New York: Basic Books.
- . 1973b. ‘Thick Description: Toward an Interpretive Theory of Culture’. In *The Interpretation of Cultures: Selected Essays*, 3–30. New York: Basic Books.
- . 1983. ‘Introduction’. In *Local Knowledge: Further Essays in Interpretive Anthropology*, 3–16. New York: Basic Books.

- Gellner, Ernest. 1984. 'The Gospel According to Ludwig'. *The American Scholar* 53 (2): 243–244, 250–251, 254–256, 258–260, 262–263.
- George, Alexander. 2012. 'Opening the Door to Cloud-Cuckoo-Land: Hempel and Kuhn on Rationality'. *Journal for the History of Analytical Philosophy* 1 (4): 1–17.
- Geuss, Raymond. 2001. *History and Illusion in Politics*. Cambridge: Cambridge University Press.
- . 2008. *Philosophy and Real Politics*. Princeton: Princeton University Press.
- . 2020. *Who Needs a Worldview?* Cambridge, MA: Harvard University Press.
- Gibbard, Allan. 1992. 'Morality and Thick Concepts (I): Thick Concepts and Warrant for Feelings'. *Proceedings of the Aristotelian Society, Supplementary Volumes* 66 (1): 267–83.
- Gibbons, Adam F. manuscript. 'Bad Language Makes Good Politics'.
- Ginsborg, Hannah. 2018. 'Normativity and Concepts'. In *The Oxford Handbook of Reasons and Normativity*. Edited by Daniel Star, 989–1014. New York: Oxford University Press.
- Glasgow, Joshua. 2020. 'Conceptual Revolution'. In *Shifting Concepts: The Philosophy and Psychology of Conceptual Variation*. Edited by Teresa Marques and Åsa Wikforss, 149–166. Oxford: Oxford University Press.
- Glasgow, Joshua, Sally Haslanger, Chike Jeffers, and Quayshawn Spencer. 2019. *What Is Race? Four Philosophical Views*. New York: Oxford University Press.
- Glock, Hans-Johann. 1996. *A Wittgenstein Dictionary*. Oxford: Blackwell.
- . 2000. 'Animals, Thoughts and Concepts'. *Synthese* 123 (1): 35–64.
- . 2006. 'Concepts: Representations or Abilities?'. In *Content, Consciousness, and Perception: Essays in Contemporary Philosophy of Mind*. Edited by Ezio Di Nucci and Conor McHugh, 36–61. Cambridge: Cambridge Scholars Press.
- . 2009a. 'Concepts, Conceptual Schemes and Grammar'. *Philosophia* 37 (4): 653.
- . 2009b. 'Concepts: Where Subjectivism Goes Wrong'. *Philosophy* 84 (1): 5–29.
- . 2010. 'Concepts: Between the Subjective and the Objective'. In *Mind, Method, and Morality: Essays in Honour of Anthony Kenny*. Edited by J. Cottingham and P. M. S. Hacker, 306–329. Oxford: Oxford University Press.
- . 2014. 'Reasons for Action: Wittgensteinian and Davidsonian Perspectives in Historical, Meta-Philosophical and Philosophical Context'. *Nordic Wittgenstein Review* 3 (1): 7–46.
- . 2017. 'Impure Conceptual Analysis'. In *The Cambridge Companion to Philosophical Methodology*. Edited by Giuseppina D'Oro and Søren Overgaard, 77–100. Cambridge: Cambridge University Press.
- . 2020. 'Concepts and Experience: A Non-Representationalist Perspective'. In *Concepts in Thought, Action, and Emotion: New Essays*. Edited by Christoph Demmerling and Dirk Schröder, 21–41. Abingdon: Routledge.
- Goetze, Trystan S. 2018. *Conceptual Responsibility*. PhD Thesis, University of Sheffield.
- . 2019. 'Conceptual Responsibility'. *Inquiry*: 1–26.
- Goldfarb, Warren. 1997. 'Wittgenstein on Fixity of Meaning'. In *Early Analytic Philosophy: Frege, Russell, Wittgenstein*. Edited by William Walker Tait, 75–89. Chicago and La Salle, Illinois: Open Court.
- Goldie, Peter. 2009. 'Thick Concepts and Emotion'. In *Reading Bernard Williams*. Edited by Daniel Callcut, 94–109. London: Routledge.
- Goldman, Alvin I. 1976. 'Discrimination and Perceptual Knowledge'. *The Journal of Philosophy* 64 (12): 771–91.

- Goodman, Nelson. 1977. *The Structure of Appearance*. 3rd ed. Dordrecht: Springer.
- . 1978. *Ways of Worldmaking*. Indianapolis: Hackett.
- . 1983. *Fact, Fiction, and Forecast*. 4 ed. Cambridge, MA: Harvard University Press.
- Gordon, Adam, René Rohrbeck, and Jan Oliver Schwarz. 2019. ‘Escaping the “Faster Horses” Trap: Bridging Strategic Foresight and Design-Based Innovation’. *Technology Innovation Management Review* 9 (8): 30–42.
- Green, Leslie. 1988. *The Authority of the State*. Oxford: Clarendon Press.
- Green, Matthew James, and Kees van Deemter. 2019. ‘The Elusive Benefits of Vagueness: Evidence from Experiments’. In *Vagueness and Rationality in Language Use and Cognition*. Edited by Richard Dietz, 63–86. Cham: Springer.
- Greene, Jamal. 2021. *How Rights Went Wrong: Why Our Obsession with Rights Is Tearing America Apart*. Boston: Houghton Mifflin Harcourt.
- Greenough, Patrick. 2020. ‘Neutralism and Conceptual Engineering’. In *Conceptual Engineering and Conceptual Ethics*. Edited by Alexis Burgess, Herman Cappelen and David Plunkett, 205–229. Oxford: Oxford University Press.
- Greenspan, Patricia. 2003. ‘The Problem with Manipulation’. *American Philosophical Quarterly* 40 (2): 155–164.
- Grönert, Peter. 2016. *Thick Concepts and Reasons for Actions*. Habilitation, University of Leipzig.
- Guest, Stephen. 2013. *Ronald Dworkin*. 3rd ed. Stanford: Stanford University Press.
- Guignon, Charles B. 2004. *On Being Authentic*. London: Routledge.
- Gustafsson, Martin. 2020. ‘Wittgenstein on Using Language and Playing Chess: The Breakdown of an Analogy and Its Consequences’. In *The Logical Alien*. Edited by Sofia Miguens, 202–221. Cambridge, MA: Harvard University Press.
- Gutzmann, Daniel. 2013. ‘Expressives and Beyond: An Introduction to Varieties of Conventional Non-Truth-Conditional Meaning’. In *Beyond Expressives: Explorations in Use-Conditional Meaning*. Edited by Daniel Gutzmann and Hans-Martin Gärtner, 1–58. Leiden: Brill.
- Hacker, Peter Michael Stephan. 2013. ‘Wittgenstein’s Anthropological and Ethnological Approach’. In *Wittgenstein: Comparisons and Context*, 111–127. Oxford: Oxford University Press.
- Hall, Edward. 2014. ‘Contingency, Confidence, and Liberalism in the Political Thought of Bernard Williams’. *Social Theory and Practice* 40 (4): 545–569.
- . 2017. ‘How To Do Realistic Political Theory (and Why You Might Want To)’. *European Journal of Political Theory* 16 (3): 283–303.
- . 2020. *Value, Conflict, and Order: Berlin, Hampshire, Williams, and the Realist Revival in Political Theory*. Chicago: University of Chicago Press.
- Hämäläinen, Nora. 2009. ‘Is Moral Theory Harmful in Practice?—Relocating Anti-theory in Contemporary Ethics’. *Ethical Theory and Moral Practice* 12 (5): 539–553.
- Hampshire, Stuart. 1983. *Morality and Conflict*. Oxford: Blackwell.
- Hampton, Jean E. 1998a. ‘The Anatomy of a Reason’. In *The Authority of Reason*. Edited by Richard Healey, 44–82. Cambridge: Cambridge University Press.
- . 1998b. ‘Reasons’ Authority’. In *The Authority of Reason*. Edited by Richard Healey, 83–122. Cambridge: Cambridge University Press.
- Handelman, Sapir. 2009. *Thought Manipulation: The Use and Abuse of Psychological Trickery*. Santa Barbara: Praeger Publishers.
- Hanna, Robert. 2015. *Cognition, Content, and the A Priori: A Study in the Philosophy of Mind and Knowledge*. Oxford: Oxford University Press.

- Hannon, Michael. 2013. 'The Practical Origins of Epistemic Contextualism'. *Erkenntnis* 78 (4): 899–919.
- . 2015. 'The Universal Core of Knowledge'. *Synthese* 192 (3): 769–786.
- . 2019. *What's the Point of Knowledge? A Function-First Epistemology*. Oxford: Oxford University Press.
- Harcourt, Edward. Manuscript. 'Consequentialism, Moralism, and the 'Administered World".
- Hardimon, Michael O. 2017. *Rethinking Race: The Case for Deflationary Realism*. Cambridge, MA: Harvard University Press.
- Harman, Gilbert. 1984. 'Logic and Reasoning'. *Synthese* 60 (1): 107–127.
- Harré, Rom. 1964. 'Concepts and Criteria'. *Mind* LXXIII (291): 353–363.
- Hart, Herbert Lionel Adolphus. 1963. 'Acts of Will and Legal Responsibility'. In *Freedom and the Will*. Edited by David Pears, 38–47. New York: Palgrave.
- . 2008a. 'Acts of Will and Responsibility'. In *Punishment and Responsibility: Essays in the Philosophy of Law*. 2nd ed, 90–112. Oxford: Oxford University Press.
- . 2008b. 'Legal Responsibility and Excuses'. In *Punishment and Responsibility: Essays in the Philosophy of Law*. 2nd ed, 28–53. Oxford: Oxford University Press.
- . 2008c. 'Negligence, Mens Rea, and Criminal Responsibility'. In *Punishment and Responsibility: Essays in the Philosophy of Law*. 2nd ed, 136–157. Oxford: Oxford University Press.
- . 2008d. 'Prolegomenon to the Principles of Punishment'. In *Punishment and Responsibility: Essays in the Philosophy of Law*. 2nd ed, 1–27. Oxford: Oxford University Press.
- . 2012. *The Concept of Law*. 3rd ed. Oxford: Oxford University Press.
- Hartman, Robert J. 2019. 'Kant Does Not Deny Resultant Moral Luck'. *Midwest Studies in Philosophy* 43 (1): 136–50.
- Hasan, Ali, and Richard Fumerton. 2018. 'Foundationalist Theories of Epistemic Justification'. In *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Fall 2018 ed.
- Haslanger, Sally. 2012. *Resisting Reality: Social Construction and Social Critique*. Oxford: Oxford University Press.
- . 2014. 'Social Meaning and Philosophical Method'. *Proceedings and Addresses of the American Philosophical Association* 88: 16–37.
- . 2018. 'Cognition as a Social Skill'. *Australasian Philosophical Review* 2 (4).
- . 2020a. 'Going On, Not in the Same Way'. In *Conceptual Engineering and Conceptual Ethics*. Edited by Alexis Burgess, Herman Cappelen and David Plunkett, 230–260. Oxford: Oxford University Press.
- . 2020b. 'How Not to Change the Subject'. In *Shifting Concepts: The Philosophy and Psychology of Conceptual Variation*. Edited by Teresa Marques and Åsa Wikforss, 235–259. Oxford: Oxford University Press.
- . 2021. *Ideology in Practice: What Does Ideology Do?* Milwaukee, Wisconsin: Marquette University Press.
- Hayward, Max Khan. 2019. 'Immoral Realism'. *Philosophical Studies* 176 (4): 897–914.
- Heal, Jane. 2007. 'Back to the Rough Ground!' Wittgensteinian Reflections on Rationality and Reason'. *Ratio* 20 (4): 403–421.
- Hegel, Georg Wilhelm Friedrich. 1975. *Aesthetics: Lectures on Fine Art*. Oxford: Oxford University Press.
- Heinze, Richard. 1925. 'Auctoritas'. *Hermes* 60 (3): 348–366.

- Heney, Diana B. 2016. *Toward a Pragmatist Metaethics*. New York: Routledge.
- Henne, Céline. 2022. *Framed and Framing Inquiry: Development and Defence of John Dewey's Theory of Knowledge*. PhD, University of Cambridge.
- Herder, Johann Gottfried. 2004. *Herder: Philosophical Writings*. Edited by Michael N. Forster. Cambridge: Cambridge University Press.
- Herman, Edward S., and Noam Chomsky. 1988. *Manufacturing Consent: The Political Economy of the Mass Media*. New York: Pantheon Books.
- Herodotus. 1920. *Histories*. Translated by A. D. Godley. Cambridge, MA: Harvard University Press.
- Heuer, Ulrike, and Gerald Lang. 2012. *Luck, Value, and Commitment: Themes from the Ethics of Bernard Williams*. Oxford: Oxford University Press.
- Heyd, David. 1997. 'Moral And Legal Luck: Kant's Reconciliation With Practical Contingency'. *Annual Review of Law and Ethics* 5: 27–42.
- Hippo, Augustine of. 1992. *Confessions*. Edited by James J. O'Donnell. Vol. 1. Oxford: Oxford University Press.
- Hirsch, Eli. 1993. *Dividing Reality*. New York: Oxford University Press.
- . 2013. 'The Metaphysically Best Language'. *Philosophy and Phenomenological Research* 87 (3): 709–716.
- Hlobil, Ulf. 2015. 'Anti-Normativism Evaluated'. *International Journal of Philosophical Studies* 23 (3): 376–395.
- Hofstadter, Richard. 2008. *The Paranoid Style in American Politics and Other Essays*. New York: Vintage.
- Horkheimer, Max. 1987. *Studien über Autorität und Familie: Forschungsberichte aus dem Institut für Sozialforschung*. Lüneburg: zu Klampen.
- Hornsby, Jennifer. 1980. *Actions*. London: Routledge.
- Horty, John F. 2012. *Reasons as Defaults*. New York: Oxford University Press.
- Hoyningen-Huene, Paul. 2013. *Systematicity: The Nature of Science*. New York: Oxford University Press.
- Huddleston, Andrew. 2019. *Nietzsche on the Decadence and Flourishing of Culture*. Oxford: Oxford University Press.
- Hume, David. 1998. *An Enquiry Concerning the Principles of Morals*. Edited by Tom L. Beauchamp. Oxford: Clarendon Press.
- . 2000. *A Treatise of Human Nature*. Edited by David Fate Norton and Mary J. Norton. Oxford: Oxford University Press.
- Hyde, Harford Montgomery. 1973. *The Trials of Oscar Wilde*. New York: Dover.
- Hyman, John. 2011. 'Action and the Will'. In *The Oxford Handbook of Wittgenstein*. Edited by Oskari Kuusela and Marie McGinn, 451–471. Oxford: Oxford University Press.
- . 2015. *Action, Knowledge, and Will*. Oxford: Oxford University Press.
- Ichikawa, Jonathan Jenkins. 2017. *Contextualising Knowledge: Epistemology and Semantics*. Oxford: Oxford University Press.
- Isaac, Manuel Gustavo. 2021. 'Post-Truth Conceptual Engineering'. *Inquiry*: 1–16.
- Jaeggi, Rahel. 2016. *Critique of Forms of Life*. Cambridge, MA: Harvard University Press.
- James, William. 1975–88. *The Works of William James*. Edited by F. H. Burkhardt, F. Bowers and I. K. Skrupskelis. 18 vols. Cambridge, MA: Harvard University Press.
- . 1981. *The Principles of Psychology, Volume II*. Edited by Frederick H. Burkhardt, Fredson Bowers and Ignas K. Skrupskelis. Cambridge, MA: Harvard University Press.

- Jamieson, Dale. 2007. 'When Utilitarians Should Be Virtue Theorists'. *Utilitas* 19 (2): 160–183.
- . 2014. *Reason in a Dark Time: Why the Struggle Against Climate Change Failed—and What It Means for Our Future*. New York: Oxford University Press.
- Jayasekera, Marie. 2016. 'Responsibility in Descartes's Theory of Judgment'. *Ergo* 3 (12): 321–47.
- Johnston, Mark. 2001. 'The Authority of Affect'. *Philosophy and Phenomenological Research* 63 (1): 181–214.
- Jorem, Sigurd. 2022. 'The Good, the Bad and the Insignificant—Assessing Concept Functions for Conceptual Engineering'. *Synthese* 200 (106): 1–20.
- Jorem, Sigurd, and Guido Löhr. 2022. 'Inferentialist Conceptual Engineering'. *Inquiry*: 1–22.
- Jouvenel, Bertrand de. 1963a. *The Pure Theory of Politics*. New Haven: Yale University Press.
- . 1963b. *Sovereignty: An Inquiry into the Political Good*. Cambridge: Cambridge University Press.
- Kaeslin, Isabel. Manuscript. 'Hermeneutic Attention'.
- Kagan, Shelly. 1989. *The Limits of Morality*. Oxford: Oxford University Press.
- Kant, Immanuel. 1900–. *Kants gesammelte Schriften*. Edited by Preussische Akademie der Wissenschaften (vols. 1–22), Deutsche Akademie der Wissenschaften zu Berlin (vol. 23) and Akademie der Wissenschaften zu Göttingen (from vol. 24). Akademieausgabe. Berlin: De Gruyter.
- . 2015. *Critique of Practical Reason*. Translated by Mary Gregor. Cambridge: Cambridge University Press.
- Kappel, Klemens. 2010. 'On Saying That Someone Knows: Themes From Craig'. In *Social Epistemology*. Edited by Adrian Haddock, Alan Millar and Duncan Pritchard, 69–88. Oxford: Oxford University Press.
- Keefe, Rosanna. 2000. *Theories of Vagueness*. Cambridge: Cambridge University Press.
- Keil, F. C. 1989. *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.
- Kekes, John. 1993. *The Morality of Pluralism*. Princeton: Princeton University Press.
- Kelp, Christoph. 2011. 'What's the Point of 'Knowledge' Anyway?'. *Episteme* 8 (1): 53–66.
- Kenny, Anthony John Patrick. 1972. 'Descartes on the Will'. In *Cartesian Studies*. Edited by R. J. Buter, 1–31. Oxford: Blackwell.
- . 1992. *The Metaphysics of Mind*. Oxford: Oxford University Press.
- Kirchin, Simon. 2010. 'The Shapelessness Hypothesis'. *Philosophers' Imprint* 10 (4): 1–28.
- Kirchner, Alexander S. 2022. *Legitimate Opposition*. New Haven: Yale University Press.
- Kitcher, Philip. 2001. *Science, Truth, and Democracy*. New York: Oxford University Press.
- . 2008. 'Carnap and the Caterpillar'. *Philosophical Topics* 36 (1): 111–127.
- . 2011. *The Ethical Project*. Cambridge, MA: Harvard University Press.
- Kletzer, Christoph, and Massimo Renzo. 2020. 'Authority and Legitimacy'. In *Cambridge Companion to the Philosophy of Law*. Edited by John Tasioulas, 191–207. Cambridge: Cambridge University Press.
- Kojève, Alexandre. 2014. *The Notion of Authority (A Brief Presentation)*. Translated by Hager Weslati. Edited by François Terré. London: Verso.

- Kolodny, Niko, and John Brunero. 2020. 'Instrumental Rationality'. In *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Spring 2020 ed.
- Koreň, Ladislav. 2021. *Practices of Reason: Fusing the Inferentialist and Scientific Image*. New York: Routledge.
- Korsgaard, Christine. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Krebs, Angelika. 1999. *Ethics of Nature: A Map*. Berlin: De Gruyter.
- Kripke, Saul A. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- . 1982. *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard University Press.
- Krisch, Nico. 2010. *Beyond Constitutionalism: The Pluralist Structure of Postnational Law*. Oxford: Oxford University Press.
- Krishnan, Nikhil, and Matthieu Queloz. 2022. 'The Shaken Realist: Bernard Williams, the War, and Philosophy as Cultural Critique'. *European Journal of Philosophy Early View*: 1–22.
- Kuhn, Thomas. 1970. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Kukla, Rebecca. 2000. 'Myth, Memory and Misrecognition in Sellars' "Empiricism and the Philosophy of Mind"'. *Philosophical Studies* 101 (2/3): 161–211.
- Kukla, Rebecca, and Mark Lance. 2009. *'Yo!' and 'Lo!': The Pragmatic Topography of the Space of Reasons*. Cambridge, MA: Harvard University Press.
- Kusch, Martin. 2006. *A Sceptical Guide to Meaning and Rules: Defending Kripke's Wittgenstein*. Chesham: Acumen.
- . 2009. 'Testimony and the Value of Knowledge'. In *Epistemic Value*. Edited by Adrian Haddock, Alan Millar and Duncan Pritchard, 60–94. Oxford: Oxford University Press.
- . 2011. 'Knowledge and Certainties in the Epistemic State of Nature'. *Episteme* 8 (1): 6–23.
- . 2013. 'Naturalized Epistemology and the Genealogy of Knowledge'. In *Contemporary Perspectives on Early Modern Philosophy: Nature and Norms in Thought*. Edited by Martin Lenz and Anik Waldow, 87–100. Dordrecht: Springer.
- . 2015. 'A Branch of Human Natural History': Wittgenstein's Reflections on Metrology'. In *Standardization in Measurement: Philosophical, Historical and Sociological Issues*. Edited by Lara Huber and Oliver Schlaudt, 11–24. London: Pickering and Chatto.
- La Boétie, Etienne de. 2016. *Discours de la servitude volontaire*. Paris: Flammarion.
- Ladd, John. 1957. *The Structure of a Moral Code: A Philosophical Analysis of Ethical Discourse Applied to the Ethics of the Navaho Indians*. Cambridge, MA: Harvard University Press.
- Laden, Anthony Simon. 2012. *Reasoning: A Social Picture*. Oxford: Oxford University Press.
- Lane, Melissa. 2018. 'Placing Plato in the History of Liberty'. *History of European Ideas* 44 (6): 702–718.
- Lanius, David. 2019. *Strategic Indeterminacy in the Law*. New York: Oxford University Press.
- . 2021. 'What Is the Value of Vagueness?'. *Theoria* 87 (3): 752–780.
- Laplace, Pierre Simon. 1951. *A Philosophical Essay on Probabilities*. Translated by F.W. Truscott and F.L. Emory. New York: Dover.

- Latham, Andrew J., Kristie Miller, and James Norton. 2019. 'Philosophical Methodology and Conceptions of Evil Action'. *Metaphilosophy* 50 (3): 296–315.
- Lau, Ting Cho. 2022. 'How to Choose Normative Concepts'. *Analytic Philosophy* Early View: 1–17.
- Lauener, Henri. 2001. 'Ethik des methodologischen Humanismus: Kritische Bemerkungen zur Relativität von Normen und zum Pluralismus von Systemen der Moral'. *Erkenntnis* 54 (1): 77–103.
- Lawlor, Krista. 2013. *Assurance: An Austinian View of Knowledge and Knowledge Claims*. Oxford: Oxford University Press.
- Lear, Jonathan. 2011. 'Foreword to the Routledge Classics Edition'. In *Ethics and the Limits of Philosophy*, ix–xiv. London: Routledge.
- Leiter, Brian. 2015. *Nietzsche on Morality*. 2 ed. London: Routledge.
- . 2019a. 'The Innocence of Becoming: Nietzsche against Guilt'. *Inquiry* 62 (1): 70–92.
- . 2019b. *Moral Psychology with Nietzsche*. Oxford: Oxford University Press.
- . 2022. 'Williams's Debt to Nietzsche: Real or Illusory?'. In *Morality and Agency: Themes from Bernard Williams*. Edited by András Szigeti and Matthew Talbert, 17–35. New York: Oxford University Press.
- Leitgeb, Hannes, and André W. Carus. 2021. 'Rudolf Carnap'. In *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Summer 2021 ed.
- Leuenberger, Muriel. 2021. 'What is the Point of Being Your True Self? A Genealogy of Essentialist Authenticity'. *Philosophy*: 1–23.
- Levy, Neil. 2011. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford: Oxford University Press.
- Lewis, David. 1973. *Counterfactuals*. Oxford: Blackwell.
- . 1983a. 'Languages and Language'. In *Philosophical Papers: Volume 1*, 163–188. Oxford: Oxford University Press.
- . 1983b. 'New Work for a Theory of Universals'. *Australasian Journal of Philosophy* 61 (4): 343–377.
- . 1983c. *Philosophical Papers: Volume 1*. Oxford: Oxford University Press.
- . 1984. 'Putnam's Paradox'. *Australasian Journal of Philosophy* 62 (3): 221–36.
- Lilla, Mark, Ronald Dworkin, and Robert Silvers. 2001. 'Introduction'. In *The Legacy of Isaiah Berlin*. Edited by Mark Lilla, Ronald Dworkin and Robert Silvers, ix–xiii. New York: New York Review of Books.
- Lindauer, Matthew. 2020. 'Conceptual Engineering as Concept Preservation'. *Ratio*: 1–8.
- Lindholm, Charles. 2013. 'The Rise of Expressive Authenticity'. *Anthropological Quarterly* 86 (2): 361–95.
- Lipscomb, Benjamin J. B. 2021. *The Women Are Up to Something: How Elizabeth Anscombe, Philippa Foot, Mary Midgley, and Iris Murdoch Revolutionized Ethics*. Oxford: Oxford University Press.
- Lloyd, Dan Edward. 1989. *Simple Minds*. Cambridge, MA: MIT Press.
- Louden, Robert B. 2007. 'The Critique of the Morality System'. In *Bernard Williams*. Edited by Alan Thomas, 104–134. Cambridge: Cambridge University Press.
- Lovibond, Sabina. 2015. 'Ethical Upbringing: From Connivance to Cognition'. In *Essays on Ethics and Feminism*, 128–145. Oxford: Oxford University Press.

- Łukomska, Agata. 2022. 'Confidence: On the Possibility of Ethical Knowledge'. In *Morality and Agency: Themes from Bernard Williams*. Edited by András Szigeti and Matthew Talbert, 110–31. New York: Oxford University Press.
- Machery, Edouard. 2017. *Philosophy Within Its Proper Bounds*. Oxford: Oxford University Press.
- MacIntyre, Alasdair C. 1978. *Against the Self-Images of the Age*. Notre Dame: University of Notre Dame Press.
- . 1988. *Whose Justice? Which Rationality?* Notre Dame: University of Notre Dame Press.
- . 2007. *After Virtue: A Study in Moral Theory*. 3rd ed. Notre Dame: University of Notre Dame Press.
- MacLeod, Alan, ed. 2019. *Propaganda in the Information Age: Still Manufacturing Consent*. Abingdon: Routledge.
- Macnish, Kevin, and Jai Galliott, eds. 2020. *Big Data and Democracy*. Edinburgh: Edinburgh University Press.
- Magee, Bryan, and Bernard Williams. 1971. 'Conversation with Bernard Williams: Philosophy and Morals'. In *Modern British Philosophy*. Edited by Bryan Magee, 150–165. New York: St. Martin's Press.
- Maitland, Frederic William. 1898. 'A Prologue to a History of English Law'. *Law Quarterly Review* 14: 13–33.
- Mameli, Matteo. 2001. 'Mindreading, Mindshaping, and Evolution'. *Biology and Philosophy* 16 (5): 595–626.
- Marconi, Diego. 1997. *Lexical Competence*. Cambridge, MA: MIT Press.
- Marcuse, Herbert. 2002. *One-Dimensional Man: Studies in the Ideology of Advanced Industrial Society*. London: Routledge.
- Marques, Teresa. 2020. 'Amelioration vs Perversion'. In *Shifting Concepts: The Philosophy and Psychology of Conceptual Variation*. Edited by Teresa Marques and Åsa Wikforss, 260–284. Oxford: Oxford University Press.
- Mashaw, Jerry L. 1983. *Bureaucratic Justice: Managing Social Security Disability Claims*. New Haven: Yale University Press.
- Matravers, Matt. 2007. *Responsibility and Justice*. London: Polity.
- Matthews, Robert J. 2010. *The Measure of Mind: Propositional Attitudes and their Attribution*. Oxford: Oxford University Press.
- McDowell, John. 1996. *Mind and World*. Cambridge, MA: Harvard University Press.
- McDowell, John Henry. 1998a. 'Aesthetic Value, Objectivity, and the Fabric of the World'. In *Mind, Value, and Reality*, 112–130. Cambridge, MA: Harvard University Press.
- . 1998b. 'Non-Cognitivism and Rule-Following'. In *Mind, Value, and Reality*, 198–220. Cambridge, MA: Harvard University Press.
- McGrath, Matthew. 2015. 'Two Purposes of Knowledge-Attribution and the Contextualism Debate'. In *Epistemic Evaluation: Purposeful Epistemology*. Edited by John Greco and David Henderson, 138–157. Oxford: Oxford University Press.
- McPherson, Tristram. 2011. 'Against Quietist Normative Realism'. *Philosophical Studies* 154 (2): 223–240.
- . 2018. 'Authoritatively Normative Concepts'. In *Oxford Studies in Metaethics Vol. 13*. Edited by Russ Shafer-Landau, 253–77. Oxford: Oxford University Press.
- . 2020a. 'Deliberative Authority and Representational Determinacy: A Challenge for the Normative Realist'. *Ergo* 6 (45): 1331–1358.

- . 2020b. *Epistemology and Methodology in Ethics*. Cambridge: Cambridge University Press.
- McPherson, Tristram, and David Plunkett. 2020. ‘Conceptual Ethics and the Methodology of Normative Inquiry’. In *Conceptual Engineering and Conceptual Ethics*. Edited by Alexis Burgess, Herman Cappelen and David Plunkett, 274–303. Oxford: Oxford University Press.
- . 2021. ‘Evaluation Turned on Itself: The Vindictory Circularity Challenge to the Conceptual Ethics of Normativity’. In *Oxford Studies in Metaethics Volume 16*. Edited by Russ Shafer-Landau, 207–32. Oxford: Oxford University Press.
- Mele, Alfred R. 2017. *Aspects of Agency: Decisions, Abilities, Explanations, and Free Will*. New York: Oxford University Press.
- Merrill, Sarah Bishop. 1998. *Defining Personhood: Toward the Ethics of Quality in Clinical Care*. Amsterdam: Rodopi.
- Midgley, Mary. 1996. ‘Philosophical Plumbing’. In *Utopias, Dolphins and Computers: Problems of Philosophical Plumbing*, 1–12. London: Routledge.
- Mill, John Stuart. 1874. *Autobiography*. London: Longmans, Green, Reader, and Dyer.
- . 1988. *The Subjection of Women*. Edited by Susan Moller Okin. Indianapolis: Hackett.
- . 2003. *On Liberty*. Edited by Mary Warnock. 2 ed. Oxford: Wiley-Blackwell.
- Miller, Alexander, and Crispin Wright, eds. 2002. *Rule-Following and Meaning*. Chesham: Acumen.
- Miller, Geoffrey. 2000. *The Matting Mind: How Sexual Choice Shaped Human Nature*. New York: Anchor Books.
- Miller, Kristie. 2010. ‘On the Concept of Sexual Perversion’. *The Philosophical Quarterly* 60 (241): 808–830.
- Millikan, Ruth Garrett. 2017. *Beyond Concepts: Unicepts, Language, and Natural Information*. Oxford: Oxford University Press.
- Mills, Claudia. 1995. ‘Politics and Manipulation’. *Social Theory and Practice* 21 (1): 97–112.
- Mommsen, Theodor. 1888. *Römisches Staatsrecht*. Leipzig: S. Hirzel.
- Montaigne, Michel de. 1967. *Essays*. Paris: Éditions du Seuil.
- Moore, Adrian W. 1991. ‘Can Reflection Destroy Knowledge?’. *Ratio* 4 (2): 97–106.
- . 1993. ‘Ineffability and Reflection: An Outline of the Concept of Knowledge’. *European Journal of Philosophy* 1 (3): 285–308.
- . 1997. *Points of View*. Oxford: Oxford University Press.
- . 2003. ‘Williams on Ethics, Knowledge and Reflection’. *Philosophy* 78: 337–354.
- . 2006a. ‘Bernard Williams: Ethics and the Limits of Philosophy’. In *Central Works of Philosophy Volume 5: The Twentieth Century: Quine and After*. Edited by John Shand. Vol. 5, 207–26. Chesham: Acumen.
- . 2006b. ‘Maxims and Thick Ethical Concepts’. *Ratio* 19: 129–147.
- . 2019a. ‘Apperception and the Unreality of Tense’. In *Language, World, and Limits: Essays in the Philosophy of Language and Metaphysics*, 143–157. Oxford: Oxford University Press.
- . 2019b. ‘How Significant is the Use/Mention Distinction?’. In *Language, World, and Limits: Essays in the Philosophy of Language and Metaphysics*, 11–38. Oxford: Oxford University Press.
- . 2019c. ‘The Metaphysics of Perspective: Tense and Colour’. In *Language, World, and Limits: Essays in the Philosophy of Language and Metaphysics*, 158–164. Oxford: Oxford University Press.

- . 2020. ‘The Concern With Truth, Sense, et al—Androcentric or Anthropocentric?’. *Angelaki* 25 (1-2): 126–134.
- . Forthcoming. ‘The Possibility of Absolute Representations’. In *Reading Rödl on Self-Consciousness and Objectivity*. Edited by James Conant and Jesse Mulder. Cambridge, MA: Harvard University Press.
- Moore, Michael S. 2010. *Act and Crime: The Philosophy of Action and Its Implications for Criminal Law*. Oxford: Oxford University Press.
- Moravcsik, Julius. 1976. ‘Ancient and Modern Conceptions of Health and Medicine’. *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* 1 (4): 337–348.
- Morton, Adam. 1980. *Frames of Mind: Constraints on the Common-Sense Conception of the Mental*. Oxford: Clarendon Press.
- Mouffe, Chantal. 2000. *The Democratic Paradox*. London: Verso.
- Moyn, Samuel. 2010. *The Last Utopia*. Cambridge, MA: Harvard University Press.
- Mühlebach, Deborah. 2019. *The Politics of Meaning: A Non-Ideal Approach to Verbal Derogation*. PhD Thesis, University of Basel.
- . 2021. ‘Semantic Contestations and the Meaning of Politically Significant Terms’. *Inquiry* 64 (8): 788–817.
- . 2022. ‘Tackling Verbal Derogation: Linguistic Meaning, Social Meaning, and Constructive Contestation’. In *The Political Turn in Analytic Philosophy: Reflections on Social Injustice and Oppressions*. Edited by D. Bordonaba Plou, V. Fernández Castro and J. R. Torices Vidal, 175–198. Berlin: De Gruyter.
- . Forthcoming. ‘Neopragmatist Inferentialism and the Meaning of Derogatory Terms – A Defence’. *Dialectica*.
- . manuscript. ‘What Are Slurs All About? On Social Structures, Linguistic Content, and the Communicative Functions of Slurs’.
- Mulhall, Stephen. 2021. *The Ascetic Ideal*. Oxford: Oxford University Press.
- Müller, Andreas. 2019. ‘Reasoning and Normative Beliefs: Not Too Sophisticated’. *Philosophical Explorations* 22 (1): 2–15.
- . 2020. *Constructing Practical Reasons*. Oxford: Oxford University Press.
- Murdoch, Iris. 1956. ‘Vision and Choice in Morality’. *Proceedings of the Aristotelian Society, Supplementary Volumes* 30: 32–58.
- . 1961. ‘Against Dryness: A Polemical Sketch’. *Encounter* 16 (1): 16–20.
- . 1999. ‘Metaphysics and Ethics’. In *Existentialists and Mystics: Writings on Philosophy and Literature*. Edited by Peter Conradi, 59–75. London: Penguin.
- . 2013. *The Sovereignty of Good*. London: Routledge.
- Nado, Jennifer. 2020. ‘Taking Control: Conceptual Engineering Without (Much) Metasemantics’. *Inquiry*: 1–27.
- . 2021a. ‘Classification Procedures as the Targets of Conceptual engineering’. *Philosophy and Phenomenological Research*.
- . 2021b. ‘Conceptual Engineering, Truth, and Efficacy’. *Synthese* 198 (7): 1507–1527.
- Nagel, Thomas. 2001. ‘Pluralism and Coherence’. In *The Legacy of Isaiah Berlin*. Edited by Mark Lilla, Ronald Dworkin and Robert Silvers, 105–111. New York: New York Review of Books.
- Nemitz, Paul. 2018. ‘Constitutional Democracy and Technology in the Age of Artificial Intelligence’. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376 (2133): 1–14.

- Ng, Geraldine. 2019. 'The Irrelativism of Distance'. In *Ethics Beyond the Limits: New Essays on Bernard Williams' Ethics and the Limits of Philosophy*. Edited by Sophie Grace Chappell and Marcel van Ackeren, 148–67. London: Routledge.
- Ng, Sai Ying. 2024. 'Relational Normativity: Williams's Thick Ethical Concepts in Confucian Ethical Communities'. *Philosophy East and West* 74 (1).
- Nguyen, C. Thi. 2020. *Games: Agency As Art*. New York: Oxford University Press.
- . 2021. 'The Seductions of Clarity'. *Royal Institute of Philosophy Supplement* 89: 227–255.
- Nietzsche, Friedrich. 1986. *Human, All Too Human*. Translated by R. J. Hollingdale. Edited by Richard Schacht. Cambridge: Cambridge University Press.
- . 1998. *On the Genealogy of Morality*. Translated by Maudemarie Clark and Alan J. Swensen. Indianapolis: Hackett.
- . 2001. *The Gay Science: With a Prelude in German Rhymes and an Appendix of Songs*. Translated by Josefine Nauckhoff and Adrian Del Caro. Edited by Bernard Williams. Cambridge: Cambridge University Press.
- . 2005a. 'Nietzsche Contra Wagner'. In *The Anti-Christ, Ecce Homo, Twilight of the Idols, and Other Writings*. Edited by Aaron Ridley and Judith Norman, 263–282. Cambridge: Cambridge University Press.
- . 2005b. *Nietzsche: The Anti-Christ, Ecce Homo, Twilight of the Idols: And Other Writings*. Translated by Judith Norman. Cambridge: Cambridge University Press.
- . 2005c. *Twilight of the Idols*. Translated by Judith Norman. Edited by Aaron Ridley and Judith Norman. Cambridge: Cambridge University Press.
- . 2009a. *Digital Critical Edition of the Complete Works and Letters*. Edited by Paolo D'Iorio. Berlin: De Gruyter. Based on the critical text by G. Colli and M. Montinari. Berlin: De Gruyter. www.nietzschesource.org/eKGWB/.
- . 2009b. 'On Truth and Lie in an Extra-Moral Sense'. In *Writings from the Early Notebooks*. Translated by Ladislaus Löb. Edited by Raymond Geuss and Alexander Nehamas, 253–64. Cambridge: Cambridge University Press.
- Nisbett, Richard. 2015. *Mindware: Tools for Smart Thinking*. London: Allen Lane.
- Nisbett, Richard, and Dov Cohen. 1996. *Culture Of Honor: the Psychology Of Violence In The South*. Boulder: Westview Press.
- Noggle, Robert. 1996. 'Manipulative Actions: A Conceptual and Moral Analysis'. *American Philosophical Quarterly* 33 (1): 43–55.
- . 2022. 'The Ethics of Manipulation'. In *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Summer 2022 edition ed.
- Norris, Andrew. 2009. 'La chaîne des raisons a une fin.' Wittgenstein et Oakeshott sur le rationalisme et la pratique'. *Cités: Philosophie, Politique, Histoire* 38: 95–108.
- Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- Nussbaum, Martha C. 1988. 'Non-Relative Virtues: An Aristotelian Approach'. *Midwest Studies in Philosophy* 13 (1): 32–53.
- . 2000. 'Why Practice Needs Ethical Theory - Particularism, Principle, and Bad Behavior'. In *The Path of the Law and Its Influence: The Legacy of Oliver Wendell Holmes*. Edited by Steven J. Burton, 50–86. Cambridge: Cambridge University Press.
- . 2001. *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*. Rev. ed. Cambridge: Cambridge University Press.
- Nyíri, J. C. 1976. 'Wittgenstein's New Traditionalism'. *Acta Philosophica Fennica* 28: 501–12.

- . 1982. 'Wittgenstein's Later Work in Relation to Conservatism'. In *Wittgenstein and his Times*. Edited by Brian McGuinness, 44–68. Chicago: University of Chicago Press.
- O'Doherty, Marianne, and Felicitas Schmieder. 2015. 'Introduction: Travels and Mobilities in the Middle Ages: From the Atlantic to the Black Sea'. In *Travels and Mobilities in the Middle Ages: From the Atlantic to the Black Sea*. Edited by Marianne O'Doherty and Felicitas Schmieder, ix–xlivi. Turnhout: Brepols.
- O'Neill, Onora. 1987. 'Abstraction, Idealization and Ideology in Ethics'. *Royal Institute of Philosophy Lecture Series* 22: 55–69.
- O'Shaughnessy, Brian. 1973. 'Trying (as the Mental 'Pineal Gland')'. *The Journal of Philosophy* 70 (13): 365–86.
- . 2008a. *The Will: A Dual Aspect Theory*. Vol. I. Cambridge: Cambridge University Press.
- . 2008b. *The Will: A Dual Aspect Theory*. Vol. II. Cambridge: Cambridge University Press.
- Ober, Josiah. 2022. *The Greeks and the Rational: The Discovery of Practical Reason*. Oakland, CA: University of California Press.
- Olsson, Erik. 2017. 'Coherentist Theories of Epistemic Justification'. In *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Spring 2017 ed.
- Orwell, George. 2008. 'Politics and the English Language'. In *All Art Is Propaganda: Critical Essays*. Edited by George Packer and Keith Gessen, 270–286. Boston and New York: Mariner.
- Owen, David. 2007. *Nietzsche's Genealogy of Morality*. Stocksfield: Acumen.
- . 2018. 'Nietzsche's Antichristian Ethics: Renaissance Virtù and the Project of Reevaluation'. In *Nietzsche and The Antichrist: Religion, Politics, and Culture in Late Modernity*. Edited by Daniel Conway, 67–88. London: Bloomsbury Academic.
- Page, Benjamin I. 1996. *Who Deliberates? Mass Media in Modern Democracy*. Chicago: University of Chicago Press.
- Pahuja, Sundhya. 2011. *Decolonising International Law: Development, Economic Growth and the Politics of Universality*. Cambridge: Cambridge University Press.
- Paine, Thomas. 1998. *Rights of Man, Common Sense, and Other Political Writings*. New York: Oxford University Press.
- Paul, Elliot Samuel. 2020. 'Cartesian Clarity'. *Philosophers' Imprint* 20 (19): 1–28.
- Paul, Laurie Ann. 2012. 'Metaphysics as Modeling: The Handmaiden's Tale'. *Philosophical Studies* 160 (1): 1–29.
- Peirce, Charles Sanders. 1931. *Collected Papers of Charles Sanders Peirce*. Edited by C. Hartshorne, P. Weiss and A. Burks. Cambridge, MA: Belknap Press.
- Percival, Philip. 1994. 'Absolute Truth'. *Proceedings of the Aristotelian Society* 94 (1): 189–214.
- Pereboom, Derk. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.
- Peregrin, Jaroslav. 2014. *Inferentialism: Why Rules Matter*. New York: Palgrave.
- Pérez Carballo, Alejandro. 2020. 'Conceptual Evaluation: Epistemic'. In *Conceptual Engineering and Conceptual Ethics*. Edited by Alexis Burgess, Herman Cappelen and David Plunkett, 304–32. Oxford: Oxford University Press.
- Perry, Stephen. 2005. 'Law and Obligation'. *The American Journal of Jurisprudence* 50 (1): 263–295.

- Pettit, Philip. 1996. 'Functional Explanation and Virtual Selection'. *British Journal for the Philosophy of Science* 47 (2): 291–302.
- . 1997. *Republicanism: A Theory of Freedom and Government*. Oxford: Oxford University Press.
- . 2008. *Made With Words: Hobbes on Language, Mind, and Politics*. Princeton: Princeton University Press.
- . 2018. *The Birth of Ethics: Reconstructing the Role and Nature of Morality*. Edited by Kinch Hoekstra. Oxford: Oxford University Press.
- . 2020. 'Analyzing Concepts and Allocating Referents'. In *Conceptual Engineering and Conceptual Ethics*. Edited by Alexis Burgess, Herman Cappelen and David Plunkett, 333–357. Oxford: Oxford University Press.
- . forthcoming. *When Minds Speak: The Social Practice that Enables Humanity*. Oxford: Oxford University Press.
- Pinder, Mark. 2019. 'Scharp on Inconsistent Concepts and their Engineered Replacements, or: Can We Mend These Broken Things?'. *Inquiry*: 1-22.
- . 2022. 'What Ought a Fruitful Explicatum to Be?'. *Erkenntnis* 87 (2): 913–932.
- Pleasants, Nigel. 1999. *Wittgenstein and the Idea of a Critical Social Theory: A Critique of Giddens, Habermas and Bhaskar*. London: Routledge. doi:10.4324/9780203056059.
- . 2002. 'Towards a Critical Use of Marx and Wittgenstein'. In *Marx and Wittgenstein: Knowledge, Morality and Politics*. Edited by Gavin Kitching and Nigel Pleasants, 160–81. London: Routledge.
- Plotica, Luke Philip. 2015. *Michael Oakeshott and the Conversation of Modern Political Thought*. Albany, NY: State University of New York Press.
- Plunkett, David. 2015. 'Which Concepts Should We Use? Metalinguistic Negotiations and The Methodology of Philosophy'. *Inquiry* 58 (7–8): 828–874.
- . 2016. 'Conceptual History, Conceptual Ethics, and the Aims of Inquiry: A Framework for Thinking about the Relevance of the History/Genealogy of Concepts to Normative Inquiry'. *Ergo: An Open Acces Journal of Philosophy* 3 (2): 27–64.
- . 2020. 'Normative Roles, Conceptual Variance, and Ardent Realism about Normativity'. *Inquiry* 63 (5): 509–534.
- Plunkett, David, and Tim Sundell. 2019. 'Metalinguistic Negotiation and Speaker Error'. *Inquiry*: 1–26.
- Plunkett, David, and Timothy Sundell. 2013a. 'Disagreement and the Semantics of Normative and Evaluative Terms'. *Philosophers' Imprint* 13 (23): 1–37.
- . 2013b. 'Dworkin's Interpretivism and the Pragmatics of Legal Disputes'. *Legal Theory* 19 (3): 242–281.
- Podosky, Paul-Mikhail Catapang. 2021. 'Agency, Power, and Injustice in Metalinguistic Disagreement'. *The Philosophical Quarterly* 72 (2): 441–464.
- Pollock, John. 1984. *The Foundations of Philosophical Semantics*. Princeton: Princeton University Press.
- Prescott-Couch, Alexander. 2015. 'Genealogy and the Structure of Interpretation'. *Journal of Nietzsche Studies* 46 (2): 239–247.
- . 2021. 'Deliberation through Misrepresentation? Inchoate Speech and the Division of Interpretive Labor*'. *Journal of Political Philosophy* 29 (4): 496–518.
- . manuscript. 'Nietzsche and the Significance of Genealogy'.
- Price, Huw. 2011. *Naturalism Without Mirrors*. Oxford: Oxford University Press.
- . 2018. 'Carnapian Voluntarism and Global Expressivism: Reply to Carus'. *The Monist* 101 (4): 468-474.

- Price, Huw, Simon Blackburn, Robert Brandom, Paul Horwich, and Michael Williams. 2013. *Expressivism, Pragmatism and Representationalism*. Cambridge: Cambridge University Press.
- Price, Huw, and David Macarthur. 2007. ‘Pragmatism, Quasi-realism and the Global Challenge’. In *New Pragmatists*. Edited by Cheryl Misak, 91–120. Oxford: Oxford University Press.
- Priest, Graham. 2006. *In Contradiction: A Study of the Transconsistent*. Oxford: Oxford University Press.
- . 2014. ‘Contradictory Concepts’. In *Contradictions: Logic, History, Actuality*. Edited by Elena Ficara, 13–26. Berlin: De Gruyter.
- . 2016. ‘Logical Disputes and the a priori’. *Logique et Analyse* 236: 347–66.
- Pritchard, Duncan. 2012. ‘The Genealogy of the Concept of Knowledge and Anti-Luck Virtue Epistemology’. In *Conceptions of Knowledge*. Edited by Stefan Tolksdorf, 159–178. Berlin: De Gruyter.
- Putnam, Hilary. 1973. ‘Meaning and Reference’. *The Journal of Philosophy* 70 (19): 699–711.
- . 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.
- . 1992. *Renewing Philosophy*. Cambridge, MA: Harvard University Press.
- . 2001. ‘Reply to Bernard Williams’ ‘Philosophy as a Humanistic Discipline’. *Philosophy* 76 (4): 605–614.
- . 2002. *The Collapse of the Fact-Value Distinction and Other Essays*. Cambridge, MA.: Harvard University Press.
- Queloz, Matthieu. 2016. ‘Wittgenstein on the Chain of Reasons’. *Wittgenstein-Studien* 7 (1): 105–130.
- . 2017. ‘Two Orders of Things: Wittgenstein on Reasons and Causes’. *Philosophy* 92 (3): 369–397.
- . 2018. ‘Williams’s Pragmatic Genealogy and Self-Effacing Functionality’. *Philosophers’ Imprint* 18 (17): 1–20.
- . 2021a. ‘Choosing Values? Williams contra Nietzsche’. *The Philosophical Quarterly* 71 (2): 286–307.
- . 2021b. *The Practical Origins of Ideas: Genealogy as Conceptual Reverse-Engineering*. Oxford: Oxford University Press.
- . 2021c. ‘The Self-Effacing Functionality of Blame’. *Philosophical Studies* 178 (4): 1361–79.
- . 2022a. ‘Genealogy, Evaluation, and Engineering’. *The Monist* 105 (4): 435–451.
- . 2022b. ‘A Shelter from Luck: The Morality System Reconstructed’. In *Morality and Agency: Themes from Bernard Williams*. Edited by András Szigeti and Matthew Talbert, 182–209. New York: Oxford University Press.
- . 2023. ‘Nietzsche’s Conceptual Ethics’. *Inquiry Published Online*: 1–31.
- Queloz, Matthieu, and Friedemann Bieber. 2022. ‘Conceptual Engineering and the Politics of Implementation’. *Pacific Philosophical Quarterly* 103 (3): 670–691.
- Queloz, Matthieu, and Damian Cueni. 2019. ‘Nietzsche as a Critic of Genealogical Debunking: Making Room for Naturalism Without Subversion’. *The Monist* 102 (3): 277–297.
- . 2021. ‘Left Wittgensteinianism’. *European Journal of Philosophy* 29 (4): 758–777.
- Queloz, Matthieu, and Marcel van Ackeren. manuscript. ‘Virtue Ethics and the Morality System’.

- Quine, Willard Van Orman. 1960. 'Carnap and Logical Truth'. *Synthese* 12 (4): 350–374.
- Radzik, Linda. 2000. 'Incorrigible Norms: Foundationalist Theories of Normative Authority'. *The Southern Journal of Philosophy* 38 (4): 633–49.
- Rathgeb, Nicole. 2020. *Die Begriffsanalyse im 21. Jahrhundert: Eine Verteidigung gegen zeitgenössische Einwände*. Paderborn: Mentis.
- Rawls, John. 1955. 'Two Concepts of Rules'. *The Philosophical Review* 64 (1): 3–32.
- . 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- . 1993. *Political Liberalism*. New York: Columbia University Press.
- Raz, Joseph. 1979. *The Authority of Law: Essays on Law and Morality*. Oxford: Clarendon Press.
- . 1986. *The Morality of Freedom*. Oxford: Clarendon Press.
- . 1995. *Ethics in the Public Domain: Essays in the Morality of Law and Politics*. Rev. ed. Oxford: Clarendon Press.
- . 1999. *Practical Reason and Norms*. Oxford: Oxford University Press.
- . 2009. *Between Authority and Interpretation: On the Theory of Law and Practical Reason*. New York: Oxford University Press.
- Rechnitzer, Tanja. 2022. *Applying Reflective Equilibrium: Towards the Justification of a Precautionary Principle*. Cham: Springer.
- Reck, Erich. 2012. 'Carnapian Explication: A Case Study and Critique'. In *Carnap's Ideal of Explication and Naturalism*. Edited by P. Wagner, 96–116. New York: Palgrave.
- . forthcoming. 'Carnapian Explication: Origins and Shifting Goals'. In *Interpreting Carnap: Critical Essays*. Edited by A. Richardson and A. T. Tuboly. Cambridge: Cambridge University Press.
- Rée, Jonathan. 1998. 'Strenuous Unbelief'. *London Review of Books*, October 15, 1998, 7–11.
- Reginster, Bernard. 2021. *The Will to Nothingness: An Essay on Nietzsche's On the Genealogy of Morality*. Oxford: Oxford University Press.
- Rescher, Nicholas. 1979. *Cognitive Systematization: A Systems Theoretic Approach to a Coherentist Theory of Knowledge*. Oxford: Blackwell.
- . 2005. *Cognitive Harmony: The Role of Systemic Harmony in the Constitution of Knowledge*. Pittsburgh, PA: University of Pittsburgh Press.
- Reynolds, Steven L. 2017. *Knowledge as Acceptable Testimony*. Cambridge: Cambridge University Press.
- Richard, Mark. 2008. *When Truth Gives Out*. New York: Oxford University Press.
- Richardson, John. 2020. *Nietzsche's Values*. New York: Oxford University Press.
- Ridley, Aaron. 2005. 'Nietzsche and the Re-Evaluation of Values'. *Proceedings of the Aristotelian Society* 105 (1): 155–175.
- Rieland, Indrek. 2022. 'Meaning Change'. *Analytic Philosophy*.
- Riggs, Jared. 2021. 'Deflating the functional turn in conceptual engineering'. *Synthese* 199 (3): 11555–11586.
- Rini, Regina. 2019. 'Epoch Relativism and Our Moral Hopelessness'. In *Ethics Beyond the Limits: New Essays on Bernard Williams' Ethics and the Limits of Philosophy*. Edited by Sophie Grace Chappell and Marcel van Ackeren, 168–87. London: Routledge.
- Ritschl, Otto. 1906. *System und systematische Methode in der Geschichte des wissenschaftlichen Sprachgebrauchs und der philosophischen Methodologie*. Bonn: C. Georgi.
- Roberts, Debbie. 2011. 'Shapelessness and the Thick'. *Ethics* 121 (3): 489–520.

- . 2013. ‘Thick Concepts’. *Philosophy Compass* 8 (8): 677–688.
- Robson, Jon, and Neil Sinclair. 2022. ‘Speculative Aesthetic Expressivism’. *The British Journal of Aesthetics*.
- Rödl, Sebastian. 2018. *Self-Consciousness and Objectivity: An Introduction to Absolute Idealism*. Cambridge, MA: Harvard University Press.
- Rorty, Richard. 1983. ‘Postmodernist Bourgeois Liberalism’. *Journal of Philosophy* 80 (10): 583–89.
- . 1989. *Contingency, Irony, and Solidarity*. Cambridge: Cambridge University Press.
- . 2021. *Pragmatism as Anti-Authoritarianism*. Edited by Eduardo Mendieta. Cambridge, MA: Belknap Press.
- Rosen, Gideon. 2022. ‘Moral Realism with a Human Face: Objectivity in Ethics and the Limits of Philosophy’. In *Agency, Fate, and Luck: Themes from Bernard Williams*. Edited by Andras Szigeti and Matthew Talbert, 132–57. New York: Oxford University Press.
- Rosen, Michael. 2013. *On Voluntary Servitude: False Consciousness and The Theory of Ideology*. Wiley.
- Ross, W. D. 1930. *The Right and the Good*. Oxford: Clarendon Press.
- Rothstein, Edward. 1998. ‘Fresh Debates On the Legacy Of Isaiah Berlin’. *The New York Times*, November 14, 9.
- Rouse, Joseph. 2015. *Articulating the World: Conceptual Understanding and the Scientific Image*. Chicago: University of Chicago Press.
- Rousseau, Jean-Jacques. 1979. *Emile, or On Education*. Translated by Alan Bloom. New York: Basic Books.
- Rudy-Hiller, Fernando. 2018. ‘The Epistemic Condition for Moral Responsibility’. In *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Fall 2018 ed.
- Russell, Bertrand. 1921. *The Analysis of Mind*. London: George Allen and Unwin.
- Russell, Paul. 2013. ‘Responsibility, Naturalism, and ‘The Morality System’’. In *Oxford Studies in Agency and Responsibility, Volume 1*. Edited by David Shoemaker, 184–204. Oxford: Oxford University Press.
- . 2017a. ‘Free Will Pessimism’. In *The Limits of Free Will: Selected Essays*, 243–276. Oxford: Oxford University Press.
- . 2017b. ‘Free Will Pessimism’. In *Oxford Studies in Agency and Responsibility*. Edited by David Shoemaker. Vol. 4, 93–120. Oxford: Oxford University Press.
- . 2017c. *The Limits of Free Will: Selected Essays*. Oxford: Oxford University Press.
- . 2018. ‘Bernard Williams: Ethics from a Human Point of View’. *Times Literary Supplement*, December 18.
- . 2019. ‘Hume’s Optimism and Williams’s Pessimism: From ‘Science of Man’ to Genealogical Critique’. In *Ethics Beyond the Limits: New Essays on Bernard Williams’ Ethics and the Limits of Philosophy*. Edited by Sophie Grace Chappell and Marcel van Ackeren, 37–52. London: Routledge.
- . 2022. ‘Free Will and the Tragic Predicament: Making Sense of Williams’. In *Morality and Agency: Themes from Bernard Williams*. Edited by András Szigeti and Matthew Talbert, 161–181. New York: Oxford University Press.
- . forthcoming. ‘Responsibility After “Morality”: Strawson’s Naturalism and Williams’ Genealogy’. In *P. F. Strawson and His Legacy*. Edited by A. Bengtson, B. De Mesel and S. Heyndels. Oxford: Oxford University Press.
- Ryle, Gilbert. 2009a. *The Concept of Mind*. Edited by Julia Tanney. London: Routledge.

- . 2009b. ‘Thinking and Reflecting’. In *Collected Papers, Volume 2: Collected Essays 1929–1968*, 479–493. Abingdon: Routledge.
- . 2009c. ‘The Thinking of Thoughts: What is ‘Le Penseur’ Doing?’. In *Collected Papers, Volume 2: Collected Essays 1929–1968*, 494–510. Abingdon: Routledge.
- Rysiew, Patrick. 2012. ‘Epistemic Scorekeeping’. In *Knowledge Ascriptions*. Edited by Jessica Brown and Mikkel Gerken, 270–294. Oxford: Oxford University Press.
- Sagar, Paul. 2017. ‘Beyond Sympathy: Smith’s Rejection of Hume’s Moral Theory’. *British Journal for the History of Philosophy* 25 (4): 681–705.
- . 2018. *The Opinion of Mankind: Sociability and the Theory of the State from Hobbes to Smith*. Princeton: Princeton University Press.
- . forthcoming. *Basic Equality*. Princeton: Princeton University Press.
- Sainsbury, R. M., and Michael Tye. 2012. *Seven Puzzles of Thought: An Originalist Theory of Concepts*. Oxford: Oxford University Press.
- Sandel, Michael. 1981. *Liberalism and the Limits of Justice*. 2 ed. Cambridge: Cambridge University Press.
- . 1996. *Democracy’s Discontent*. 2 ed. Cambridge, MA: Harvard University Press.
- Santarelli, Matteo. 2022. ‘Improving Concepts, Reshaping Values: Pragmatism and Ameliorative Projects’. *Inquiry*: 1–19.
- Santelli, Mauro. 2020. ‘Redescribing Final Vocabularies: A Rortian Picture of Identity and Selfhood’. *European Journal of Pragmatism and American Philosophy* XII (1): 1–21.
- Saunders, Kevin W. 1988. ‘Voluntary Acts and the Criminal Law: Justifying Culpability Based on the Existence of Volition’. *University of Pittsburgh Law Review* 49 (2): 443–76.
- Sawyer, Sarah. 2018. ‘The Importance of Concepts’. *Proceedings of the Aristotelian Society* 118 (2): 127–147.
- . 2020a. ‘The Role of Concepts in Fixing Language’. *Canadian Journal of Philosophy* 50 (5): 555–65.
- . 2020b. ‘Talk and Thought’. In *Conceptual Engineering and Conceptual Ethics*. Edited by Alexis Burgess, Herman Cappelen and David Plunkett, 379–395. Oxford: Oxford University Press.
- . 2020c. ‘Truth and Objectivity in Conceptual Engineering’. *Inquiry* 63 (9–10): 1001–1022.
- Scanlon, Thomas M. 1992. ‘The Aims and Authority of Moral Theory’. *Oxford Journal of Legal Studies* 12 (1): 1–23.
- . 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- . 2003. ‘Thickness and Theory’. *The Journal of Philosophy* 100 (6): 275–287.
- Schaffer, Jonathan. 2004. ‘From Contextualism to Contrastivism’. *Philosophical Studies* 119 (1–2): 73–103.
- Scharp, Kevin. 2013. *Replacing Truth*. Oxford: Oxford University Press.
- . 2020. ‘Philosophy as the Study of Defective Concepts’. In *Conceptual Engineering and Conceptual Ethics*. Edited by Alexis Burgess, Herman Cappelen and David Plunkett, 396–416. Oxford: Oxford University Press.
- . 2021. ‘Conceptual Engineering for Truth: Aletheic Properties and New Aletheic Concepts’. *Synthese* 198 (Suppl 2): 647–688.
- Scheffler, Samuel. 1992. *Human Morality*. Oxford: Oxford University Press.
- Schiappa, Edward. 2003. *Defining Reality: Definitions and the Politics of Meaning*. Carbondale: Southern Illinois University Press.

- Schneck, Ariane Cäcilie. 2019. 'Elisabeth of Bohemia's Neo-Peripatetic Account of the Emotions'. *British Journal for the History of Philosophy* 27 (4): 753–770.
- Schroeter, Laura, and François Schroeter. 2015. 'Rationalizing Self-Interpretation'. In *The Palgrave Handbook of Philosophical Methods*. Edited by Chris Daly, 419–447. New York: Palgrave.
- Searle, John. 2010. *Making the Social World: The Structure of Human Civilization*. Oxford: Oxford University Press.
- Sederberg, Peter C. 1984. *The Politics of Meaning: Power and Explanation in the Construction of Social Reality*. Tucson: University of Arizona Press.
- Sellars, Wilfrid. 1958. 'Counterfactuals, Dispositions, and the Causal Modalities'. In *Minnesota Studies in the Philosophy of Science*. Edited by Herbert Feigl, Michael Scriven and Grover Maxwell. Vol. II, 225–308. Minneapolis: University of Minnesota Press.
- . 1997. *Empiricism and the Philosophy of Mind*. Edited by Richard Rorty. Cambridge, MA: Harvard University Press.
- Shackelford, Todd K. 2005. 'An Evolutionary Psychological Perspective on Cultures of Honor'. *Evolutionary Psychology* 3 (1): 381–391.
- Shields, Matthew. 2021a. 'Conceptual Change and Future Paths for Pragmatism'. *The Southern Journal of Philosophy* 59 (3): 405–434.
- . 2021b. 'Conceptual Domination'. *Synthese* 199 (5): 15043–15067.
- . 2021c. 'On Stipulation'. *European Journal of Philosophy* 29 (4): 1100–1114.
- Shun, Kwong-loi, and David B. Wong, eds. 2004. *Confucian Ethics: A Comparative Study of Self, Autonomy, and Community*. New York: Cambridge University Press.
- Sider, Theodore. 2011. *Writing the Book of the World*. Oxford: Oxford University Press.
- Simion, Mona. 2018. 'The 'Should' in Conceptual Engineering'. *Inquiry* 61 (8): 914–928.
- Simion, Mona, and Christoph Kelp. 2020. 'Conceptual Innovation, Function First'. *Noûs* 54 (4): 985–1002.
- Sinclair, Neil. 2018. 'Conceptual Role Semantics and the Reference of Moral Concepts'. *European Journal of Philosophy* 26 (1): 95–121.
- . 2021. *Practical Expressivism*. Oxford: Oxford University Press.
- Skinner, Quentin. 1994. 'Modernity and Disenchantment: Some Historical Reflections'. In *Philosophy in an Age of Pluralism: The Philosophy of Charles Taylor in Question*. Edited by Tully James and Daniel M. Weinstock, 37–48. Cambridge: Cambridge University Press.
- . 1997. 'Rhetoric and Conceptual Change'. In *Quentin Skinner's Rhetoric of Conceptual Change*. Edited by Kari Palonen, 60–73. London: SAGE.
- . 1998. *Liberty Before Liberalism*. Cambridge: Cambridge University Press.
- . 2002. *Visions of Politics: Volume 2: Renaissance Virtues*. Cambridge: Cambridge University Press.
- . 2009. 'A Genealogy of the Modern State'. In *Proceedings of the British Academy, Volume 162, 2008 Lectures*. Edited by Ron Johnston, 325–370. Oxford: Oxford University Press.
- . 2017. 'Machiavelli and the Misunderstanding of Princely Virtù'. In *Machiavelli on Liberty and Conflict*. Edited by David Johnston, Nadia Urbinati and Camila Vergara, 139–63. Chicago: University of Chicago Press.
- Skyrms, Brian. 1996. *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

- . 2004. *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press.
- Smith, David Livingstone. 2020. *On Inhumanity: Dehumanization and How to Resist It*. New York: Oxford University Press.
- Smith, Helen. 2021. ‘Clinical AI: Opacity, Accountability, Responsibility and Liability’. *AI and Society* 36 (2): 535–545.
- Smith, Michael. 2013. ‘On the Nature and Significance of the Distinction between Thick and Thin Ethical Concepts’. In *Thick Concepts*. Edited by Simon Kirchin, 97–120. Oxford: Oxford University Press.
- Smith, Sophie. 2021. ‘Historicizing Rawls’. *Modern Intellectual History* 18 (4): 906–939.
- Smithson, Robert. 2020. ‘Conceptual Cartography’. *Inquiry*: 1–26.
- Smyth, Nicholas. 2018. ‘Integration and Authority: Rescuing the ‘One Thought Too Many’ Problem’. *Canadian Journal of Philosophy* 48 (6): 812–830.
- . 2019. ‘The Inevitability of Inauthenticity: Bernard Williams and Practical Alienation’. In *Ethics beyond the Limits: New Essays on Bernard Williams’ Ethics and the Limits of Philosophy*. Edited by Sophie Grace Chappell and Marcel van Ackeren, 188–208. London: Routledge.
- . 2020. ‘Socratic Reductionism in Ethics’. *European Journal of Philosophy* 28 (4): 970–85.
- . 2022. ‘Nothing Personal: On the Limits of the Impersonal Temperament in Ethics’. *Journal of Value Inquiry* 56 (1): 67–83.
- Snell, Bruno. 1953. *The Discovery of the Mind: The Greek Origins of European Thought*. Translated by T. G. Rosenmeyer. Cambridge, MA: Harvard University Press.
- Soames, Scott. 2011. ‘The Value of Vagueness’. In *Philosophical Foundations of Language in the Law*. Edited by Andrei Marmor and Scott Soames, 14–30. Oxford: Oxford University Press.
- Sober, Elliott. 1984. *The Nature of Selection*. Cambridge, MA: MIT Press.
- Spencer, Jack. 2016. ‘Relativity and Degrees of Relationality’. *Philosophy and Phenomenological Research* XCII (2): 432–459.
- Srinivasan, Amia. 2015. ‘The Archimedean Urge’. *Philosophical Perspectives* 29 (1): 325–362.
- . 2019. ‘Genealogy, Epistemology and Worldmaking’. *Proceedings of the Aristotelian Society* CXIX (2): 127–156.
- . manuscript. *The Contingent World: Genealogy, Epistemology, Politics*.
- Stampe, Dennis W. 1987. ‘The Authority of Desire’. *The Philosophical Review* 96 (3): 335–381.
- Steinbeck, John. 2006. *The Grapes of Wrath*. New York: Penguin.
- Steinberger, Florian. 2016. ‘How Tolerant Can You Be? Carnap on Rationality’. *Philosophy and Phenomenological Research* 92 (3): 645–668.
- Stevenson, Charles L. 1944. *Ethics and Language*. New Haven: Yale University Press.
- Stevenson, Charles Leslie. 1938. ‘Persuasive Definitions’. *Mind* XLVII (187): 331–350.
- Stich, Stephen. 1983. *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.
- Stine, Gail C. 1976. ‘Skepticism, Relevant Alternatives, and Deductive Closure’. *Philosophical Studies* 29 (4): 249–61.
- Stocker, Michael. 1990. *Plural and Conflicting Values*. Oxford: Clarendon Press.
- Strawson, Peter Frederick. 1959. *Individuals*. London: Routledge.

- . 2008a. ‘Freedom and Resentment’. In *Freedom and Resentment and Other Essays*, 1–28. Abingdon: Routledge.
- . 2008b. ‘Social Morality and Individual Ideal’. In *Freedom and Resentment and Other Essays*, 29–49. Abingdon: Routledge.
- Strevens, Michael. 2008. *Depth: An Account of Scientific Explanation*. Cambridge, MA: Harvard University Press.
- Stroud, Sarah. 2019. ‘Conceptual Disagreement’. *American Philosophical Quarterly* 56 (1): 15–27.
- Sunstein, Cass R. 1996. *Legal Reasoning and Political Conflict*. New York: Oxford University Press.
- . 2001. *One Case at a Time: Judicial Minimalism on the Supreme Court*. Cambridge, Massachusetts: Harvard University Press.
- . 2016. *The Ethics of Influence: Government in the Age of Behavioral Science*. New York: Cambridge University Press.
- Susskind, Jamie. 2018. *Future Politics: Living Together in a World Transformed by Tech*. New York: Oxford University Press.
- Tasioulas, John. 1998. ‘Relativism, Realism, and Reflection’. *Inquiry* 41 (4): 377–410.
- Taylor, Charles. 1985. *Philosophy and the Human Sciences: Philosophical Papers*. Vol. II. Cambridge: Cambridge University Press.
- . 1989. *Sources of the Self*. Cambridge, MA: Harvard University Press.
- Teichmann, Roger. 2021. ‘Conceptual Corruption’. In *Cora Diamond on Ethics*. Edited by Maria Balaska, 33–55. New York: Palgrave.
- Temelini, Michael 2015. *Wittgenstein and the Study of Politics*. Toronto: University of Toronto Press.
- Tersman, Folke. 2018. ‘Recent Work on Reflective Equilibrium and Method in Ethics’. *Philosophy Compass* 13 (6): e12493.
- Testini, Francesco. 2020. ‘Testing Pragmatic Genealogy in Political Theory: The Curious Case of John Rawls’. *European Journal of Political Theory*.
- . 2021a. *Crabwalk: Applying Pragmatic Genealogy to Contextualist Political Theory*. PhD Thesis: University of Milan.
- . 2021b. ‘Genealogical Solutions to the Problem of Critical Distance: Political Theory, Contextualism and the Case of Punishment in Transitional Scenarios’. *Res Publica*.
- Thomas, Alan. 2006. *Value and Context: The Nature of Moral and Political Knowledge*. Oxford: Clarendon Press.
- Thomasson, Amie L. 2015. *Ontology Made Easy*. New York: Oxford University Press.
- . 2020a. *Norms and Necessity*. New York: Oxford University Press.
- . 2020b. ‘A Pragmatic Method for Normative Conceptual Work’. In *Conceptual Engineering and Conceptual Ethics*. Edited by Alexis Burgess, Herman Cappelen and David Plunkett, 435–458. Oxford: Oxford University Press.
- . 2022. ‘How Should We Think about Linguistic Function?’. *Inquiry*: 1–32.
- Thompson, Edward Palmer. 1975. *Whigs and Hunters. The Origin of the Black Act*. London: Penguin.
- Tiberius, Valerie. 2018. *Well-Being as Value Fulfillment: How We Can Help Each Other to Live Well*. Oxford: Oxford University Press.
- Tooley, Michael. 1972. ‘Abortion and Infanticide’. *Philosophy & Public Affairs* 2 (1): 37–65.
- . 1983. *Abortion and Infanticide*. Oxford: Clarendon Press.

- Toulmin, Stephen. 1953. *An Examination of the Place of Reason in Ethics*. Cambridge: Cambridge University Press.
- Trilling, Lionel. 1972. *Sincerity and Authenticity*. Cambridge, MA: Harvard University Press.
- Tye, Larry. 1998. *The Father of Spin: Edward L. Bernays and The Birth of Public Relations*. New York: Henry Holt.
- Van Deemter, Kees. 2010. *Not Exactly: In Praise of Vagueness*. Oxford: Oxford University Press.
- Väyrynen, Pekka. 2013. *The Lewd, The Rude, and the Nasty: A Study of Thick Concepts in Ethics*. Oxford: Oxford University Press.
- Vecht, Joost Jacob. 2020. ‘Open Texture Clarified’. *Inquiry*: 1–21.
- Vélez, Carissa. 2020. *Privacy is Power: Why and How You Should Take Back Control of Your Data*. London: Bantam Press.
- Veluwenkamp, Herman, and Jeroen van den Hoven. 2023. ‘Design for Values and Conceptual Engineering’. *Ethics and Information Technology* 25 (1): 1–12.
- Virgil. 2007. *Aeneid*. Oxford: Oxford University Press.
- Waismann, Friedrich. 1945. ‘Verifiability’. *Proceedings of the Aristotelian Society, Supplementary Volumes XIX*: 119–50.
- Wakil, Samantha. 2021. ‘Experimental Explications for Conceptual Engineering’. *Erkenntnis*.
- Waller, Bruce N. 2011. *Against Moral Responsibility*. Cambridge, MA: MIT Press.
- Walzer, Michael. 1983. *Spheres of Justice*. Oxford: Blackwell.
- . 1987. *Interpretation and Social Criticism*. Cambridge, MA: Harvard University Press.
- Wanderer, Jeremy. 2008. *Robert Brandom*. Montreal: McGill-Queen’s University Press.
- Weaver, Bryan R., and Kevin Scharp. 2019. *Semantics for Reasons*. Oxford: Oxford University Press.
- Weber, Marcel. 2005. *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.
- Weber, Max. 2019. *Economy and Society: A New Translation*. Edited and translated by Keith Tribe. Cambridge, MA: Harvard University Press.
- Wei, Xintong. 2022. ‘A Practice-based Account of The Truth Norm of Belief’. *Episteme*: 1–21.
- Weisberg, Michael. 2007. ‘Three Kinds of Idealization’. *The Journal of Philosophy* 104 (12): 639–659.
- . 2013. *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.
- Westermarck, Edvard. 1924. *The Origin and Development of the Moral Ideas*. 2nd ed. London: Macmillan.
- Wiggins, David. 1990. ‘Moral Cognitivism, Moral Relativism and Motivating Moral Beliefs’. *Proceedings of the Aristotelian Society* 91: 61–85.
- . 2002. ‘Claims of Need’. In *Needs, Values, Truth*. 3rd ed, 1–57. Oxford: Clarendon Press.
- Williams, Bernard. 1963. ‘Postscript’. In *Freedom and the Will*. Edited by David Pears, 105–137. New York: Palgrave.
- . 1973a. ‘Deciding to Believe’. In *Problems of the Self*, 136–51. Cambridge: Cambridge University Press.
- . 1973b. ‘Egoism and Altruism’. In *Problems of the Self*, 250–265. Cambridge: Cambridge University Press.

- . 1973c. ‘Ethical Consistency’. In *Problems of the Self*, 166–186. Cambridge: Cambridge University Press.
- . 1973d. ‘The Idea of Equality’. In *Problems of the Self*, 230–249. Cambridge: Cambridge University Press.
- . 1981a. ‘Conflicts of Values’. In *Moral Luck*, 71–82. Cambridge: Cambridge University Press.
- . 1981b. ‘Internal and External Reasons’. In *Moral Luck*, 101–113. Cambridge: Cambridge University Press.
- . 1981c. ‘Moral Luck’. In *Moral Luck*, 20–39. Cambridge: Cambridge University Press.
- . 1981d. ‘Ought and Moral Obligation’. In *Moral Luck*, 114–23. Cambridge: Cambridge University Press.
- . 1981e. ‘Persons, Character and Morality’. In *Moral Luck*, 1–19. Cambridge: Cambridge University Press.
- . 1981f. ‘The Truth in Relativism’. In *Moral Luck*, 132–143. Cambridge: Cambridge University Press.
- . 1981g. ‘Utilitarianism and Moral Self-Indulgence’. In *Moral Luck*, 40–53. Cambridge: Cambridge University Press.
- . 1984. ‘The Scientific and the Ethical’. In *Royal Institute of Philosophy Lecture Series*. Edited by Stuart C. Brown. Vol. 17, 209–228. Cambridge: Cambridge University Press.
- . 1985. ‘Theories of Social Justice—Where Next?’. In *Equality and Discrimination: Essays in Freedom and Justice*. Edited by S. Guest and A. Milne. Stuttgart: F. Steiner.
- . 1986. ‘Reply to Simon Blackburn’. *Philosophical Books* 27 (4): 203–208.
- . 1989. ‘Social Justice’. *Journal of Social Philosophy* 20 (1–2): 68–73.
- . 1993. *Shame and Necessity*. Berkeley: University of California Press.
- . 1995a. ‘Ethics’. In *Philosophy 1: A Guide Through the Subject*. Edited by A. C. Grayling, 545–582. Oxford: Oxford University Press.
- . 1995b. ‘Formal and Substantial Individualism’. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 123–34. Cambridge: Cambridge University Press.
- . 1995c. ‘How Free Does the Will Need to Be?’. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 3–21. Cambridge: Cambridge University Press.
- . 1995d. ‘Internal Reasons and the Obscurity of Blame’. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 35–45. Cambridge: Cambridge University Press.
- . 1995e. ‘Moral Luck: A Postscript’. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 241–247. Cambridge: Cambridge University Press.
- . 1995f. ‘Must a Concern for the Environment Be Centred on Human Beings?’. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 233–240. Cambridge: Cambridge University Press.
- . 1995g. ‘Nietzsche’s Minimalist Moral Psychology’. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 65–78. Cambridge: Cambridge University Press.
- . 1995h. ‘The Point of View of the Universe: Sidgwick and the Ambitions of Ethics’. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 153–71. Cambridge: Cambridge University Press.

- . 1995i. 'Replies'. In *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*. Edited by J. E. J. Altham and Ross Harrison, 185–224. Cambridge: Cambridge University Press.
- . 1995j. 'Saint-Just's Illusion'. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 135–150. Cambridge: Cambridge University Press.
- . 1995k. 'Truth in Ethics'. *Ratio* 8 (3): 227–238.
- . 1995l. 'Voluntary Acts and Responsible Agents'. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 22–34. Cambridge: Cambridge University Press.
- . 1995m. 'What Does Intuitionism Imply?'. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 182–191. Cambridge: Cambridge University Press.
- . 1995n. 'What Has Philosophy to Learn from Tort Law?'. In *The Philosophical Foundations of Tort Law*. Edited by David G. Owen, 487–498. Oxford: Oxford University Press.
- . 1995o. 'Who Needs Ethical Knowledge?'. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 203–212. Cambridge: Cambridge University Press.
- . 1996. 'Contemporary Philosophy: A Second Look'. In *The Blackwell Companion to Philosophy*. Edited by Nicholas Bunnin and Eric P. Tsui-James, 23–35. Oxford: Blackwell.
- . 1998. 'Virtues and Vices'. In *Routledge Encyclopedia of Philosophy*. Edited by Edward Craig. London: Routledge.
- . 1999. 'Seminar with Bernard Williams'. *Ethical Perspectives* 6 (3–4): 243–265.
- . 2001a. 'Liberalism and Loss'. In *The Legacy of Isaiah Berlin*. Edited by Mark Lilla, Ronald Dworkin and Robert Silvers, 91–103. New York: New York Review of Books.
- . 2001b. *Morality: An Introduction to Ethics*. Cambridge: Cambridge University Press.
- . 2002. *Truth and Truthfulness: An Essay in Genealogy*. Princeton: Princeton University Press.
- . 2003. 'Relativism, History, and the Existence of Values'. In *The Practice of Value*. Edited by R. Jay Wallace, 106–117. Oxford: Clarendon Press.
- . 2005a. 'Conflicts of Liberty and Equality'. In *In the Beginning Was the Deed: Realism and Moralism in Political Argument*. Edited by Geoffrey Hawthorne, 115–127. Princeton: Princeton University Press.
- . 2005b. *Descartes: The Project of Pure Enquiry*. London: Routledge.
- . 2005c. 'From Freedom to Liberty: The Construction of a Political Value'. In *In the Beginning Was the Deed: Realism and Moralism in Political Argument*. Edited by Geoffrey Hawthorne, 75–96. Princeton: Princeton University Press.
- . 2005d. 'Human Rights and Relativism'. In *In the Beginning Was the Deed: Realism and Moralism in Political Argument*. Edited by Geoffrey Hawthorne, 62–74. Princeton: Princeton University Press.
- . 2005e. 'In the Beginning Was the Deed'. In *In the Beginning Was the Deed: Realism and Moralism in Political Argument*. Edited by Geoffrey Hawthorne, 18–28. Princeton: Princeton University Press.
- . 2005f. *In the Beginning Was the Deed: Realism and Moralism in Political Argument*. Edited by Geoffrey Hawthorne. Princeton: Princeton University Press.
- . 2005g. 'The Liberalism of Fear'. In *In the Beginning Was the Deed: Realism and Moralism in Political Argument*. Edited by Geoffrey Hawthorne, 52–61. Princeton: Princeton University Press.

- . 2005h. ‘Modernity and the Substance of Ethical Life’. In *In the Beginning Was the Deed: Realism and Moralism in Political Argument*. Edited by Geoffrey Hawthorne, 40–51. Princeton: Princeton University Press.
- . 2005i. ‘Pluralism, Community and Left Wittgensteinianism’. In *In the Beginning Was the Deed: Realism and Moralism in Political Argument*. Edited by Geoffrey Hawthorne, 29–39. Princeton: Princeton University Press.
- . 2005j. ‘Realism and Moralism in Political Theory’. In *In the Beginning Was the Deed: Realism and Moralism in Political Argument*. Edited by Geoffrey Hawthorne, 1–17. Princeton: Princeton University Press.
- . 2006a. ‘The Actus Reus of Dr. Caligari’. In *Philosophy as a Humanistic Discipline*. Edited by A. W. Moore, 97–108. Princeton: Princeton University Press.
- . 2006b. ‘An Essay on Collingwood’. In *The Sense of the Past: Essays in the History of Philosophy*. Edited by Myles Burnyeat, 341–360. Princeton: Princeton University Press.
- . 2006c. ‘The Human Prejudice’. In *Philosophy as a Humanistic Discipline*. Edited by Adrian W. Moore, 135–154. Princeton: Princeton University Press.
- . 2006d. ‘The Legacy of Greek Philosophy’. In *The Sense of the Past: Essays in the History of Philosophy*. Edited by Myles Burnyeat, 3–48. Princeton: Princeton University Press.
- . 2006e. ‘Moral Responsibility and Political Freedom’. In *Philosophy as a Humanistic Discipline*. Edited by A. W. Moore, 119–125. Princeton: Princeton University Press.
- . 2006f. ‘Pagan Justice and Christian Love’. In *The Sense of the Past: Essays in the History of Philosophy*. Edited by Myles Burnyeat, 71–82. Princeton: Princeton University Press.
- . 2006g. ‘Philosophy as a Humanistic Discipline’. In *Philosophy as a Humanistic Discipline*. Edited by Adrian W. Moore, 180–199. Princeton: Princeton University Press.
- . 2006h. ‘Political Philosophy and the Analytical Tradition’. In *Philosophy as a Humanistic Discipline*. Edited by A. W. Moore, 155–168. Princeton: Princeton University Press.
- . 2006i. ‘Subjectivism and Toleration’. In *Philosophy as a Humanistic Discipline*. Edited by Adrian W. Moore, 86–96. Princeton: Princeton University Press.
- . 2006j. ‘Values, Reasons, and the Theory of Persuasion’. In *Philosophy as a Humanistic Discipline*. Edited by A. W. Moore, 109–118. Princeton: Princeton University Press.
- . 2009. ‘A Mistrustful Animal’. In *Conversations on Ethics*. Edited by Alex Voorhoeve, 195–214. Oxford: Oxford University Press.
- . 2010. ‘Who Needs Ethical Knowledge?’. *Royal Institute of Philosophy Supplement* 35: 213–222.
- . 2011. *Ethics and the Limits of Philosophy*. Routledge Classics. London: Routledge.
- . 2013. ‘Introduction’. In *Concepts and Categories: Philosophical Essays*. Edited by Henry Hardy. 2nd ed, xxix–xxxix. Princeton: Princeton University Press.
- . 2014a. ‘The Last Word, by Thomas Nagel’. In *Essays and Reviews 1959–2002*. Edited by Michael Woods, 371–387. Princeton: Princeton University Press.
- . 2014b. ‘A Matter of Principle, by Ronald Dworkin’. In *Essays and Reviews 1959–2002*. Edited by Michael Woods, 256–261. Princeton: Princeton University Press.
- . 2014c. ‘Realism with a Human Face, by Hilary Putnam’. In *Essays and Reviews 1959–2002*. Edited by Michael Woods, 320–326. Princeton: Princeton University Press.

- . 2014d. ‘Sense and Sensibilia and Philosophical Papers, by J. L. Austin’. In *Essays and Reviews 1959–2002*. Edited by Michael Woods, 40–45. Princeton: Princeton University Press.
- . 2019. ‘Left-Wing Wittgenstein’. *Common Knowledge* 25 (1–3): 321–31.
- . 2021. ‘Ethics, A Matter of Style? Introduction to the French Edition’. *Philosophical Inquiries* 9 (2): 269–284.
- Williams, Michael. 2010. ‘Pragmatism, Minimalism, Expressivism’. *International Journal of Philosophical Studies* 18 (3): 317–330.
- . 2013. ‘How Pragmatists can be Local Expressivists’. In *Expressivism, Pragmatism and Representationalism*. Edited by Huw Price, 128–144. Cambridge: Cambridge University Press.
- Williams, Robert. A. 2012. *Tragedy, Recognition, and the Death of God: Studies in Hegel and Nietzsche*. Oxford: Oxford University Press.
- Williamson, Timothy. 2003. ‘Understanding and Inference’. *Proceedings of the Aristotelian Society Supplementary Volume* 77 (1): 249–293.
- . 2017. ‘Model-Building in Philosophy’. In *Philosophy’s Future: The Problem of Philosophical Progress*. Edited by Russell Blackford and Damien Broderick, 159–172. Hoboken, NJ: Wiley-Blackwell.
- . 2018a. *Doing Philosophy: From Common Curiosity to Logical Reasoning*. Oxford: Oxford University Press.
- . 2018b. ‘Model-Building as a Philosophical Method’. *Phenomenology and Mind* 15: 16–22.
- . 2020. *Philosophical Method: A Very Short Introduction*. Oxford: Oxford University Press.
- Wilson, George, and Samuel Shpall. 2012. ‘Action’. In *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Winter 2016 ed.
- Winch, Peter. 1958. *The Idea of a Social Science and its Relation to Philosophy*. London: Routledge.
- Wittgenstein, Ludwig. 1958. *The Blue and Brown Books: Preliminary Studies for the Philosophical Investigations*. Oxford: Blackwell.
- . 1966. *Lectures and Conversations on Aesthetics, Psychology and Religious Beliefs: Compiled from Notes taken by Yorick Smythies, Rush Rhees and James Taylor*. Edited by C. Barrett. Oxford: Blackwell.
- . 1969. *On Certainty*. Edited by G. E. M. Anscombe and G. H. von Wright. Oxford: Blackwell.
- . 1974. *Philosophical Grammar*. Edited by R. Rhees. Oxford: Blackwell.
- . 1975. *Philosophical Remarks*. Translated by Raymond Hargreaves and Roger White. Edited by Rush Rhees. Oxford: Blackwell.
- . 1978. *Remarks on the Foundations of Mathematics*. Edited by G. H. von Wright, R. Rhees and G. E. M. Anscombe. 3rd ed. Oxford: Blackwell.
- . 1979. *Wittgenstein’s Lectures: Cambridge 1932–35. (From the notes of Alice Ambrose and Margaret Macdonald)*. Edited by Alice Ambrose. Oxford: Blackwell.
- . 1981. *Zettel*. Edited by G. E. M. Anscombe and G. H. von Wright. 2nd ed. Oxford: Blackwell.
- . 2000. *Wittgenstein’s Nachlass. The Bergen Electronic Edition*. Edited by The Wittgenstein Archives at the University of Bergen. Oxford: Oxford University Press.

- . 2005. *The Big Typescript*. Edited by C. Grant Luckhardt and Maximilian A. E. Aue. Oxford: Blackwell.
- . 2009. *Philosophische Untersuchungen = Philosophical Investigations*. Translated by G. E. M. Anscombe, P. M. S. Hacker and Joachim Schulte. Edited by P. M. S. Hacker and J. Schulte. rev. 4th ed. Chichester, U.K.: Wiley-Blackwell.
- Wodak, Daniel. 2019. ‘Mere Formalities: Fictional Normativity and Normative Authority’. *Canadian Journal of Philosophy* 49 (6): 828–850.
- Wolf, Susan. 2010. *Meaning in Life and Why It Matters*. Princeton: Princeton University Press.
- . 2015a. ‘Good-for-Nothings’. In *The Variety of Values: Essays on Morality, Meaning, and Love*, 67–85. New York: Oxford University Press.
- . 2015b. ‘Loving Attention’. In *The Variety of Values: Essays on Morality, Meaning, and Love*, 163–80. New York: Oxford University Press.
- Wolf, William Clark. 2020. ‘The Authority of Conceptual Analysis in Hegelian Ethical Life’. In *An Ethical Modernity? Hegel’s Concept of Ethical Life Today – Its Limits and Potential*. Edited by Tereza Matějčková and Jiří Chotaš, 15–35. Leiden: Brill.
- Wollstonecraft, Mary. 2014. *A Vindication of the Rights of Woman*. Edited by Eileen Hunt Botting. New Haven: Yale University Press.
- Wong, David B. 1991. ‘Is There a Distinction between Reason and Emotion in Mencius?’ *Philosophy East and West* 41 (1): 31–44.
- . 2006. *Natural Moralities: A Defense of Pluralistic Relativism*. New York: Oxford University Press.
- Wright, Crispin. 1992. *Truth and Objectivity*. Cambridge, MA: Harvard University Press.
- Yablo, Stephen. 1993. ‘Hop, Skip and Jump: The Agonistic Conception of Truth’. *Philosophical Perspectives* 7 (Language and Logic): 371–396.
- Yap, Audrey. 2022. ‘Conceptual Engineering and Neurath’s Boat: A Return to the Political Roots of Logical Empiricism’. In *The Political Turn in Analytic Philosophy: Reflections on Social Injustice and Oppression*. Edited by David Bordonaba Plou, Víctor Fernández Castro and José Ramón Torices, 31–52. Berlin: De Gruyter.
- Yeager, Daniel Brian. 2006. *J. L. Austin and the Law: Exculpation and the Explication of Responsibility*. Lewisburg: Bucknell University Press.
- Zacka, Bernardo. 2017. *When the State Meets the Street: Public Service and Moral Agency*. Cambridge, MA: Harvard University Press.
- Zagzebski, Linda Trinkaus. 1996. ‘Foreknowledge, Causal Relations, and Subjunctive Conditionals’. In *The Dilemma of Freedom and Foreknowledge*, 98–124. Oxford University Press.
- Zawidzki, Tadeusz Wiesław. 2013. *Mindshaping: A New Framework for Understanding Human Social Cognition*. Cambridge, MA: MIT Press.
- Zimmerman, Michael J. 2002. ‘Taking Luck Seriously’. *The Journal of Philosophy* 99 (11): 553–76.
- Zuboff, Shoshana. 2015. ‘Big Other: Surveillance Capitalism and the Prospects of an Information Civilization’. *Journal of Information Technology* 30 (1): 75–89.
- . 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.