



LUND UNIVERSITY

Minimal Optimism

Reading P. F. Strawson on Responsibility

Emilsson, Anton

2024

Document Version:

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Emilsson, A. (2024). *Minimal Optimism: Reading P. F. Strawson on Responsibility*. [Doctoral Thesis (monograph), Department of Philosophy]. Department of Philosophy, Lund University.

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

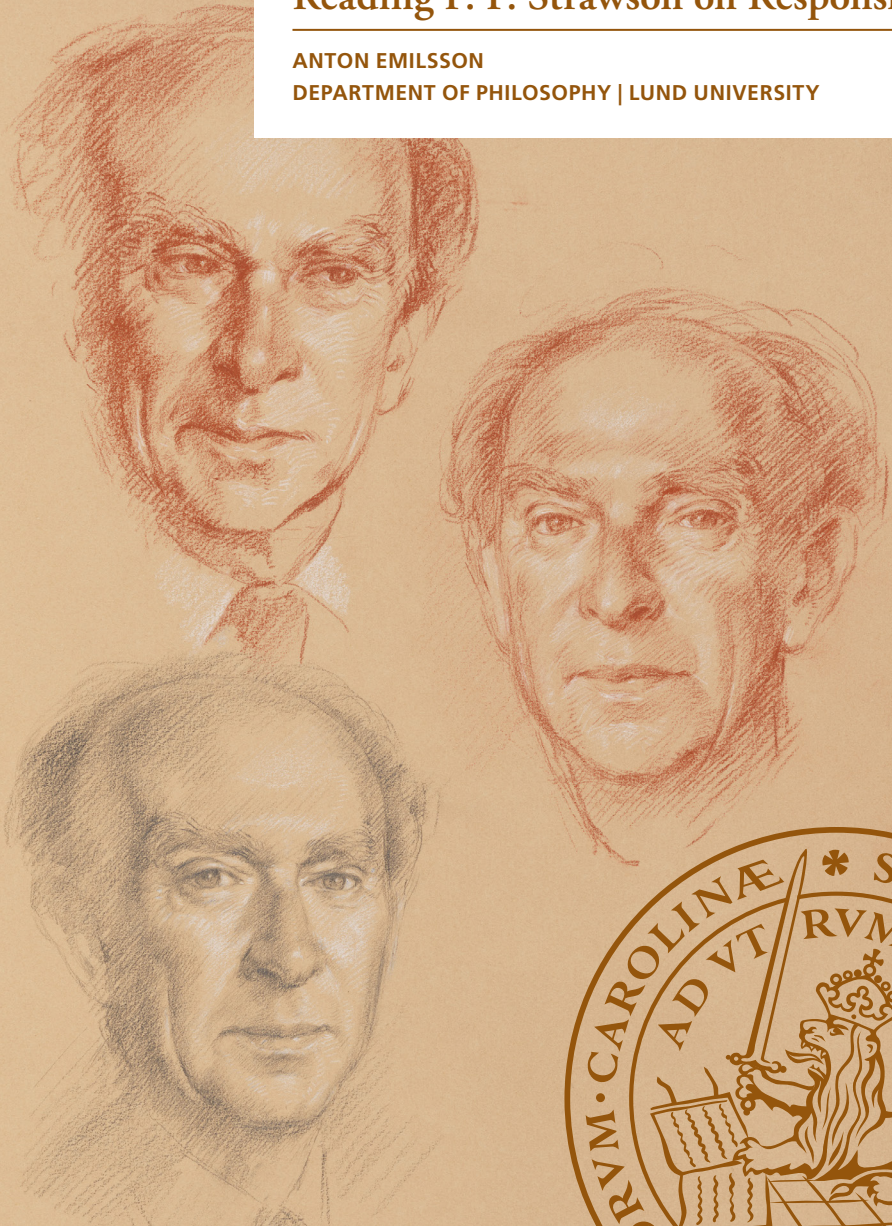
PO Box 117
221 00 Lund
+46 46-222 00 00

Minimal Optimism

Reading P. F. Strawson on Responsibility

ANTON EMILSSON

DEPARTMENT OF PHILOSOPHY | LUND UNIVERSITY





Minimal Optimism
Reading P. F. Strawson on Responsibility

Minimal Optimism

Reading P. F. Strawson on Responsibility

Anton Emilsson



LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Joint Faculties of Humanities and Theology at Lund University to be publicly defended on 6th of April at 11.00 in LUX:C216, Helgonavägen 3, Lund.

Faculty opponent
Benjamin De Mesel

Organization: LUND UNIVERSITY

Document name: Doctoral dissertation

Date of issue: 2024-04-06

Author: Anton Emilsson

Title and subtitle: Minimal Optimism: Reading P. F. Strawson on Responsibility

Abstract:

Peter Strawson's 'Freedom and Resentment' (1962) has reoriented the way moral philosophers think about responsibility. There is a general sense that Strawson's approach holds the promise that we may move beyond the traditional discussion of free will in recognition of the fact that whether determinism is true is irrelevant for whether we are responsible beings. This is *Strawson's Promise*. This thesis sets itself to investigate how we may make good on this promise.

In the first part of this thesis, three existent ways of making good on Strawson's Promise are considered: Justin Coates's transcendental argument against responsibility scepticism, Pamela Hieronymi's Statistical Reading of "the central argument", and the Reversal Move, on two different articulations. For different reasons, neither is found adequate for the task.

In the second part of this thesis, the case is made that we may make good on Strawson's Promise if we replace the Standard Reading of Strawson's Inescapability Claim—that we are naturally and inescapably committed to the concept of responsibility—with the Minimal Optimist Reading. Distinguishing between *the concept* of responsibility in its non-historical form and *any particular conception* of responsibility, Minimal Optimism is the view that we are inescapably committed to the concept of responsibility, but not so to any particular conception of responsibility. The concept of responsibility is both *conceptually* necessary and *practically* necessary. Unlike other framework commitments, however, the concept of responsibility is not essential to any experience which we can make truly intelligible to ourselves as such. In the case of responsibility, the practical necessity of the concept of responsibility is therefore in a sense more basic. A kind of naturalistic explanation—a *pragmatic genealogy*—is presented as supplementing Strawson's approach. This supports the Inescapability Claim by showing that for beings like us, already on the assumption only of maximally generic needs and situations, it is practically necessary that we have *some* conception of responsibility. Seeing that the concept of responsibility is inescapable in the sense that it is indispensable, we are vindicated in living by the concept. The truth of determinism is relevant neither for whether we have a practical need for the concept of responsibility nor for the concept of responsibility that follows simply from the relevant needs and situations.

The question *whether* we are responsible is not a question for the Minimal Optimist. In this sense, we may move beyond the traditional discussion regarding free will. Strawson's relaxed attitude to responsibility is grounded in a recognition of the fact that the concept of responsibility is indispensable for beings like us. This, however, does not straightforwardly imply any optimism about our particular conception of responsibility. Towards the end, we consider some issues regarding what Minimal Optimism implies about our particular conception of responsibility.

Key words: P. F. Strawson, responsibility, free will, scepticism, naturalism, free will scepticism, compatibilism, philosophical methodology, pragmatic genealogy, blame.

ISBN (print): 978-91-89874-21-3

Language: English

ISBN (digital): 978-91-89874-22-0

Number of pages: 217

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature

Date 2024-02-19

Minimal Optimism

Reading P. F. Strawson on Responsibility

Anton Emilsson



LUND
UNIVERSITY

Coverphoto: *Sir Peter Frederick Strawson*, by Peter Wardle. Coloured chalk, 1989. NPG 6645.
Copyright: National Portrait Gallery, London. Reproduced under Non-commercial Licence.

Copyright: Anton Emilsson

The Joint Faculties of Humanities and Theology
Department of Philosophy

ISBN (print): 978-91-89874-21-3

ISBN (digital): 978-91-89874-22-0

Printed in Sweden by Media-Tryck, Lund University
Lund 2024



Media-Tryck is a Nordic Swan Ecolabel
certified provider of printed material.
Read more about our environmental
work at www.mediatryck.lu.se

MADE IN SWEDEN 

Till Hilma och Linnea

Table of Contents

Acknowledgements	13
1. Introduction	15
I. The Promise of ‘Freedom and Resentment’	15
III. The Optimist, the Pessimist, the Sceptic, and Strawson	23
III. Overview of Chapters	26
2. Scepticism, Naturalism, and Transcendental Arguments.....	33
I. Introduction.....	33
II. Transcendental Arguments.....	36
III. A Transcendental Argument in ‘Freedom and Resentment’?	44
IV. Answering Glock’s Question	50
V. Naturalism.....	54
VI. Conclusion.....	59
3. The Incoherence Puzzle.....	61
I. Introduction.....	61
II. The Statistical Reading and the Conventional Way	62
III. The Second Incoherence Puzzle.....	67
IV. A Second Accusation of Incoherence? An Irrelevance.....	69
V. The Self-Contradiction Puzzle	76
VI. Conclusion.....	82
4. The Reversal Move	85
I. Introduction.....	85
II. The History of the Reversal Move	87
By Way of Introduction.....	87
A Possible Retraction: The Straight Path or the Short Cut?	91

III.	Two Reversal Theses	92
	The Fitting Response-Dependence Thesis.....	92
	The Meaning-Based Account	99
IV.	Strawson's Promise and the Reversal Move	104
V.	Conclusion.....	107
5.	The Inescapability Claim	109
I.	Introduction.....	109
II.	The Standard Reading.....	110
	The Standard Reading of the Inescapability Claim	110
	The Standard Criticisms of the Inescapability Claim	112
	Against the Consensus.....	115
III.	Refocusing the Inescapability Claim.....	116
IV.	Delimiting the Inescapability Claim	118
V.	The Inescapability of the Bounds of Sense	125
	Strawson's Analytic Kantianism	125
	Framework Commitments	130
	Doubly Inescapable: Two Senses of Necessity	132
VI.	A Twist or Complication.....	135
VII.	Conclusion.....	137
6.	The Unanswered Question	139
I.	Introduction.....	139
II.	The Explanatory Task	141
III.	Justification Contra Explanation	145
IV.	An Answer to the Unanswered Question	147
	The Method of Pragmatic Genealogy.....	149
	A Pragmatic Genealogy of Blame	151
	The Concept of Responsibility	155
V.	The Answer Strawson Should Have Given.....	158
VI.	Explanation, Justification, and Vindication	162
VII.	The Fundamental Irrelevance of Determinism.....	167
VIII.	Conclusion.....	168

7. Minimal Optimism & Strawson's Optimism	171
I. Introduction.....	171
II. On Inadequacy and Historical Sense.....	172
III. Minimal Optimism and Strawson's Spirit.....	177
IV. Strawson's Optimism Reconsidered	177
V. Strawson's Realism and the Limits of Description.....	186
VI. Conclusion.....	188
8. Conclusion	191
Appendix: Letter to Paul Russell, from P. F. Strawson, 1992	195
Typescript.....	197
Photocopy.....	198
References	201

Acknowledgements

I've been extremely lucky, inside and outside of the halls of academia, to have been surrounded and held up and partly formed by such wonderful, brilliant, and extraordinary people. This thesis owes so much to the care, concern, and intelligence of all of you.

I would like to begin by thanking Matthew Talbert, my main supervisor. For this thesis, for the many forks in the path not visible but part of it, and for much else beside it, you've constantly been a great support. A careful reader, a patient listener, and admirably clear and constructive in comments and discussions. I will be forever grateful.

Paul Russell, my second supervisor, thank you for teaching me about depth in philosophy. You opened up ways of thinking about what philosophy is supposed to be that have changed me. Your encouragement from the very beginning, even before, that meant so much for me. Thank you.

Matthieu Queloz, my third supervisor, thank you for a lovely time in Basel. Thank you for all your guidance and support, and not least for your simultaneously patient and piercing feedback.

I've had the fortune not only to pursue a PhD in philosophy, but to do so at the Department of Philosophy at Lund University, a truly wonderful institution. This fact owes significantly to Toni Rønnow-Rasmussen's warm and personal stewardship of our subject. I'm fortunate to have been part of this period of philosophical activity in Lund.

I would like to thank Mattias Gunnemyr, Marta Johansson Werkmäster, Jiwon Kim, Marianna Leventi, Gloria Mähringer, Martin Sjöberg, Robert Pál-Wallin, Alexander Velichkov, and Jakob Werkmäster for making B532 such a warm place.

I would also like to thank the people of and around the Lund Gothenburg Responsibility Project, which I have been fortunate enough to have been a part of. In addition to those already mentioned, thank you, András Szigeti and Björn Petersson; thank you, Olle Blomberg, Hadi Fazeli, Yulia Kanygina, Shervin MirzaeiGhazi, Daniel Telech, Patrick Todd, and Caroline Torpe Touborg; and thank you, Gunnar Björnsson, Dane Leigh Gogoshin, Per-Erik Milam, Sean Shields, and David Shoemaker.

The central idea of this thesis was first presented at a workshop in Leuven; I thank Benjamin De Mesel, Sybren Heyndels, and Stefaan Cuypers for valuable discussions during my visit at KU Leuven. I'm grateful to Maria Alvarez for serving as my mock opponent, providing very helpful comments in conversation and writing on a previous draft of this thesis.

Many more than those already mentioned have contributed with conversations concerning many ideas that have shaped this thesis in several respects; at higher and lower seminars, at conferences, over breakfast, lunch, fika and dinner, in bars and on walks, you have all contributed to a very stimulating intellectual environment. In this respect, I wish to thank Jesper Ahlin Marceta, David Alm, Henrik Andersson, Agn s Sophia Constance Baehni, Monika Betzler, Dominik Boll, Ellen Davidsson, Dan Egonsson, Alexander Edlich, Andr s Garcia, Fritz G vertsson, Hubert H gemark, Ingvar Johansson, Jenny Magnusson, Elsa Magnell, Michael McKenna, Leonard Menges, Max Minden Ribeiro, August Olsen, Wlodek Rabniowicz, Lubomira Radoilska, Signe Sav n, Jakob Stenseke, Michael Thorne, Melina Tsapos, Nicola Zetti, and Fredrik  sterblom.

I owe a special thanks to Alexander Velichkov for so many discussions of ideas that are central to this thesis and of many that are not in it thanks to him. Also Niklas Dahl, thank you for many wonderfully stimulating and explorative discussions on a wide range of topics. I would also like to extend a special thanks to Tobias Adolfsson for many years of delightful discussions.

Special thanks are also due to some who contributed with comments at a very late stage of the process: Niklas Dahl, Marianna Leventi, Robert P l-Wallin, Matthieu Queloz, Paul Russell, Martin S j berg, Matthew Talbert, Alexander Velichkov, Jakob Werkm ster, and Nicola Zetti, thank you.

This PhD would not have been possible without the administrative work of David Alm, Kristina Arnrup Thorsbro, Anna Cagnan Enh rning, Tobias Hansson Wahlberg, Petter Johansson, Martin J nsson, Tomas Persson, Anna Smedberg-Eivers, and Anna  stberg. Thank you.

Many have contributed to the completion of this project less directly. I wish to thank my parents, Anne Bengtsson and H kan Emilsson, and my brother, Andreas Emilsson, for your constant love and support. I'm also grateful for the love of so many friends and of my extended family. Last, though I cannot possibly convey how fortunate I am to have them in my life, I wish to thank Hilma and Linnea.

1. Introduction

I. The Promise of ‘Freedom and Resentment’

If you have not read Peter Strawson’s ‘Freedom and Resentment’,¹ or even have not read it in quite some time, close this book. Read it. If you cannot simply move on from that, come back.

It is a very suggestive paper. It evokes, in some, the sense that if we only carefully attend to what it is actually like to live with each other—what it is to be involved in the variety of interpersonal relationships we are, in fact, so very familiar with—then we recognise that whether the truth of the thesis of determinism is confirmed or not is completely irrelevant. Seeing this, we may finally move beyond the stultifying rehearsals of the problem of freedom which has been the preoccupation of philosophers for ages. In leaving their ethereal concerns behind, we are freed from the intractabilities of the free will debate.

While it evokes this sense with some, the paper provokes others. It provokes precisely because it is so very suggestive, first of all in the sense that it has that strong impression on some people, but also in the sense that it only *suggests* this new way forward, without dealing adequately with the many intricate complications of the issue. If the perennial problem of our freedom as agents is to be solved—indeed, *surpassed*—then this will just not do. Strawson even openly confesses to not even know what the thesis of determinism is—how then can he purport to solve the free will problem?²

Whatever one’s sentiment—whether any of these or, most plausibly, some combination of them—it is undeniable that ‘Freedom and Resentment’ has had

¹ Originally published in 1962 in *Proceedings of the British Academy* 48, 187-211. Reprinted in *Freedom and Resentment and other essays* (2008b).

² Michael McKenna (2005, p. 164) aptly remarks that Strawson’s approach “is a bit like scotch; either you love it or you hate it”. Speaking from personal experience, and as is well-known in the case of scotch, one may learn to love it.

a profound impact on moral philosophy since its publication.³ Why is it that it has been so very influential?

Paul Snowden conjectures, in his foreword to Strawson's (2008b) *Freedom and Resentment and other essays*, that the paper has been so influential because, for one thing, people just want it to be true that we are responsible—so any account with that optimistic conclusion will be attractive—and, second, a sentimentalist approach to morality, which Strawson is taken to propose, is always appealing—that is, again, any such account will be attractive. Whether or not it is in fact true that these ideas are so attractive, as an explanation of why 'Freedom and Resentment' *in particular* has been so very influential, this is not particularly illuminating.

In part, the paper's influence may be explained in broadly sociological terms. Sir Peter Frederick Strawson (1919-2006) was after all one of the most significant philosophers of the last century. (Let's take this occasion to get to know him a bit.)

He came to Oxford in 1937 to study English, but immediately changed to Philosophy, Politics, and Economics (PPE). When graduating in 1940, Strawson received only a second-class mark in philosophy, not the first-class mark that one would have expected. He himself thought that this disappointment had put his newly acquired dream of becoming a philosopher at Oxford beyond reach.⁴ However, after his service in the war, Strawson won

3 As the authors of one very recent anthology on Strawson's work say, "The influence of 'Freedom and Resentment' on contemporary discussions of moral responsibility can hardly be overestimated" (Heyndels, Bengtson, & De Mesel, 2024a, p. 5). There are two collections of papers on 'Freedom and Resentment': McKenna & Russell (2008b) and Shoemaker & Tognazzini (2014). McKenna & Pereboom (2016) dedicates a separate chapter to Strawson's approach in their introduction to free will. There is no contemporary introduction to responsibility that could omit discussing Strawson's arguments and impact. Recently, Pamela Hieronymi (2020) has published a short book on the paper. Several important works in moral philosophy and responsibility self-consciously build on themes from Strawson's approach in 'Freedom and Resentment', including Wallace (1994), Russell (1995), Darwall (2006), and McKenna (2012).

4 See Strawson (1998a, p. 5). One rumour has it that the younger examiner forgot Strawson's exam scripts in the back of a taxi and thus could not properly argue against the older, more conservative examiner (see, Gomes, 2019). John Searle reports that he confronted Isaiah Berlin (the younger examiner) about Strawson's surprising mark (See Searle's tribute in (Walker, 2008, p. 18)). Berlin told Searle that the older examiner (Sandy Lindsay) hated the new and trendy 'linguistic philosophy' and wanted to give Strawson a 'third'; it was apparently a source of great embarrassment for Berlin that he had only managed to argue Strawson up to a 'second' and not to the 'first' that he should have gotten. (It might be of note that, the year that Strawson's *Individuals* (1959) appeared, Berlin wrote in a private letter about the book, "It looks to me a decent, but ultimately provincial, performance. That is my view of him in general but I am told that I am deeply mistaken and that I might have

the prestigious John Locke prize and, having impressed Gilbert Ryle, one of the examiners, became a fellow of University College, Oxford, in 1948. He soon rose to stardom with his paper “On Referring” (1950), arguing against Bertrand Russell’s theory of definite description, and his debate against J. L. Austin on truth—which seemed to many as a case of the youngling beating the master at his own game.⁵

In 1968, he succeeded Gilbert Ryle as Waynflete Professor in Metaphysical Philosophy at Magdalen College, Oxford, a position he held until his retirement in 1987. From the 1950s onwards he had an enormous influence on philosophy in the analytic tradition. This gets some recognition by the fact that he was elected to the British Academy in 1960—upon which he gave the lecture ‘Freedom and Resentment’—and was knighted in 1977. He went on to write several books—*Introduction to Logical Theory* (1952), *Individuals: An Essay in Descriptive Metaphysics* (1959), *The Bounds of Sense: An Essay on Kant’s Critique of First Reason* (1966, reprint 2019), *Scepticism and Naturalism* (1985, reprint 2008), and *Analysis and Metaphysics* (1992)—as well as to publish many papers in the philosophy of language, metaphysics, and epistemology.⁶

Given his place in the history of analytic philosophy, it is not altogether surprising that ‘Freedom and Resentment’ has had the impact it has. But it is hardly a sufficient explanation. One of the many ironies about ‘Freedom and Resentment’⁷ is that, as the overview of his work above suggests, it was one of his very few ventures into moral philosophy.⁸ By his own word, that paper,

thought the same about Kant. But I stick to my view: it may just be his face, his manners, his general behaviour, but I cannot help it, I take the view of the romantic philosophers that personality and philosophical capacity and outlook are not totally dissociable, [...] and what is provincial about any personality remains provincial about all manifestations” (Berlin, 2009.)

5 On the exchange between Austin (1950) and Strawson (1950), see Snowdon & Gomes (2023, § 3) and Krishnan (2023, pp. 233-235).

6 Many of which are collected in his *Logico-Linguistic Papers* (1971), *Freedom and Resentment and other essays* (1974, reprint 2008), *Subject and Predicate in Logic and Grammar* (1974), *Entity and Identity and Other Essays* (1997), and *Philosophical Writings* (2011d).

7 To mention just one minor one. It is reported (Allais, 2014, n. 3) that Strawson, by his own word, wrote ‘Freedom and Resentment’ in a single draft. We might find that extraordinary. This is the paper, as Philip Pettit (2006/7) suggests, might be “the most influential philosophical paper of the twentieth century”. Then again, Strawson often felt like the writing itself came fairly easy (Pivčević, 1989, p. 4; Strawson, 1998a, p. 19).

8 Clifford Brown (2006), in his introduction to Strawson’s philosophy, does not make much—hardly anything—of ‘Freedom and Resentment’.

together with ‘Social Morality and Individual Ideal’ which was written around the same time, “effectively embody all I have thought or have to say in an area which, important as I recognise it to be, I have never found so intellectually gripping as the areas to which I have given more attention” (Strawson, 1998a, p. 11). He declared his relative lack of interest in moral philosophy more than once (e.g., Magee & Strawson, 1971, p. 126; Strawson, Ezcurdia, Sainsbury, & Davies, 2008, p. 95f.) and regularly joked that he would turn to moral philosophy only when his intellectual powers were waning (cf., Snowdon & Gomes, 2023).

This does not mean that ‘Freedom and Resentment’ is not an important paper. What it suggests, rather, is that ‘Freedom and Resentment’ may be profitably read in light of the rest of Strawson’s work; as a natural extension of his overarching philosophical program. In large part, this thesis aims to explore the manifest interconnectedness, overlap, and continuity between Strawson’s approach to responsibility and his other writings.⁹ I should immediately stress that this is a project in no way completed here; in this respect, the present thesis is incomplete and imperfect; much else remains to be learned (and unlearned) about Strawson’s approach to responsibility from further consideration of his approach to other issues in philosophy.

Another irony of the paper, related to the last point, is a discrepancy in what Strawson’s general contribution to philosophy has been taken to be and what his contribution to the philosophy of responsibility has been taken to be. With respect to responsibility, Michael McKenna and Paul Russell (2008a, p. 5) present Strawson’s approach as follows:

Strawson’s strategy in ‘Freedom and Resentment’ involves turning away from conceptual issues about the analysis of ‘freedom’ and ‘responsibility’ and taking a closer look at what actually goes on when we hold a person responsible. That is to say, his methodology depends less on conceptual analysis and more on a descriptive account of actual human psychology.

By others, however, coming from outside of the contemporary responsibility discussion, Strawson has been hailed as “a pioneer of what is generally called conceptual analysis” (Heawood, 2006, p. 25). Likewise, while Strawson’s approach to responsibility is commonly seen as eschewing or circumventing the metaphysical discussion of free will, “and to bring the discussion to a more homely level, by focusing on the way in which we see each other when we have reactive attitudes” (Allais, 2014, p. 37), his philosophy is also, from a

⁹ I am grateful, in particular, to the work of Benjamin De Mesel and Sybren Heyndels for first illustrating to me the fruitfulness of so approaching Strawson’s work on responsibility.

different perspective, regularly credited with the revival of metaphysics within the analytic tradition (Glock, 2003a, 2003b, 2012, 2017; Peter M. S. Hacker, 2001). This might have us wondering whether the influence of ‘Freedom and Resentment’ has not been somewhat skewed (more so, that is, than is perhaps inevitably the case with any influential text).

Some of this discrepancy, however, owes to terminological differences. The project of ‘descriptive metaphysics’ that Strawson outlined and pursued does not amount to the same thing as the metaphysical debate regarding the freedom of the will that he is taken to have circumvented or surpassed. Rather, Strawson’s project is a kind of second-order conceptual exercise, a matter of describing our conceptual scheme. The aim of philosophy, according to Strawson, “is to achieve a kind of reflective conceptual self-consciousness” (Pivčević, 1989, p. 4).¹⁰ It is not a matter of measuring our concepts against some true essence of reality. It requires, instead, that we consider the actual life of our concepts:

the actual use of linguistic expressions remains [the philosopher’s] sole and essential point of contact with the reality which he wishes to understand, conceptual reality; for this is the only point from which the actual mode of operation of concepts can be observed. If he severs this vital connection, all his ingenuity and imagination will not save him from lapses into the arid or the absurd. (Strawson, 2011a, p. 90)

This might just be what is meant by “a more homely level” of analysis (Allais, 2014, p. 37) or that we need to attend to the “practical role” of the concept of responsibility (cf., McKenna & Pereboom, 2016, p. 124).

Part of Strawson’s approach, however, has been misrepresented or unappreciated as it pertains to issues of responsibility. He did not turn away from conceptual analysis. He championed a form of conceptual analysis he called ‘connective’, rather than the more prevalent ‘reductive’ method. Reductive analysis seeks “to get a clear grasp of complex meanings by reducing them, without remainder, to simple meanings”; paradigmatically, by stating “the necessary and sufficient conditions for the correct application of the concept” (Strawson, 1992a, p. 17f.). Complaints of imprecision or incompleteness in Strawson’s account are sometimes just expressions of an idea of analysis which Strawson does not share, one “in which the conception of the simple, at least as the ideal limit of analysis, plays an essential part” (Strawson, 1992a, p. 18). Strawson’s ideal form of analysis is different.

¹⁰ See also Strawson (2011e, p. 225; 2019, p. 34). Sybren Heyndels (2020), in explicating Strawson’s meta-philosophy, takes this to be Strawson’s central idea.

Let us abandon the notion of perfect simplicity in concepts; let us abandon even the notion that analysis must always be in the direction of greater simplicity. Let us imagine, instead, the model of an elaborate network, a system, of connected items, concepts, such that the function of each item, each concept, could, from the philosophical point of view, be properly understood only by grasping its connections with the others, its place in the system—perhaps better still, the picture of a set of interlocking systems of such a kind. (Strawson, 1992a, p. 19)

It seems, in so far as Strawson's account in 'Freedom and Resentment' provides the impetus for various reductive analyses and even non-conceptual explanations of responsibility, that his influence on the debate, real nonetheless, has been interestingly and not insignificantly warped by a treatment of 'Freedom and Resentment' partly insensitive to Strawson's more general philosophical program.

Sometimes, however, the complaints of imprecision or incompleteness may be said to be primarily complaints over style—not that this is unconnected to the issue of method.¹¹ While such complaints are sometimes heard, Strawson's style—the elegance of his prose, how confidently he moves in this difficult terrain, the memorable phrases and sharp remarks—must be counted among the salient reasons why the paper has been so very influential.¹² Strawson's style is indeed part of what makes the paper (in more than one respect) so very suggestive. While clear and lucid, it is not a style "which aims to head off any ambiguity or any implication which some reader, perhaps a very perverse reader, might improperly take up; a style, that is to say, which seeks precision by total mind control, through issuing continuous and rigid interpretative directions", as Bernard Williams (2006a, p. 343) describes the style of the "Cook Wilsonians" or "realists" dominating Oxford a few generations before Strawson.¹³

Lucy Allais (2024) remarks that the paper is like a Rorschach test, foremost revealing of your own view of responsibility and free will, open to almost any

11 I'm personally inclined to agree with Bernard Williams that "[t]he problem of how to find a style in moral philosophy is actually one of the deepest questions about it" (Magee & Williams, 1971, p. 164).

12 As Thomas Nagel puts it in his tribute to Strawson: "his clear and beautiful writings, free of technical apparatus, provided an inspiring demonstration that rigorous philosophy could still be a form of literature" (Walker, 2008, p. 10).

13 Paul Grice once said, having then moved from Oxford to California, "If you can't put it in symbols, it isn't worth saying". When told of his former tutor's remark, Strawson, still in Oxford, smiled and said, "Oh no! If you *can* put it into symbols, then it isn't worth saying" (Peter M. S Hacker, 1996, p. 311, n. 69).

interpretation.¹⁴ This claim gathers some support from the fact that there is a great diversity among views which declare themselves “Strawsonian”. We find, for example, broadly Humean (e.g., Russell, 2017b, 2017d), Kantian (e.g., Allais, 2014; Coates, 2017), and Wittgensteinian (e.g., Bengtson, 2019; Campbell, 2017; De Mesel, 2022a) readings of ‘Freedom and Resentment’. Strawson self-consciously took influence from all three. Arguably, though, the Kantian and the Wittgensteinian influences are more pronounced; more, that is, than the general tenor of the responsibility literature might suggest. In Kant’s first *Critique*, Strawson reports having “found a depth, a range, a boldness, and a power unlike anything I had previously encountered” (2003, p. 8). The first time he read Wittgenstein, he says, “I felt that I was, for the first time, seeing thought *naked*, as it were” (Pivčević, 1989, p. 7, original emphasis). It might be worth bearing this in mind when trying to understand ‘Freedom and Resentment’.¹⁵

If one can allow it to be, ‘Freedom and Resentment’ is a fascinating, deep, subtle, and highly suggestive paper. It should however be admitted that the arguments of the paper are not altogether easy to pin down—at least not without the feeling that one has reduced them. This might indeed be “annoying” and “frustrating”, as Allais (2014, 2024) records, but it is also telling of something more. In dealing with the difficulty of recapitulating Strawson’s particular arguments, Michael McKenna and Derk Pereboom (2016, p. 124), in their authoritative introduction to free will, make the case that it is not any individual argument that makes ‘Freedom and Resentment’ so special.

The force of Strawson’s essay ‘Freedom and Resentment’ *cannot* be understood in terms of a specific line of argumentation for or against any single proposition regarding free will or moral responsibility. (McKenna & Pereboom, 2016, p. 142, original emphasis)

Rather, the significance of the paper consists in Strawson seeking to *reorient* the very way in which we approach the issue of free will and responsibility (McKenna & Pereboom, 2016, p. 143). It is in this sense, of him reorienting the discussion, that the great appeal of Strawson’s approach lies. What explains the influence of Strawson’s approach is that it awards the sense that we may

¹⁴ A sense of the text I think that many share.

¹⁵ See also Strawson (1998a, p. 20). This is not to deny that Hume might also have an important influence on Strawson; it might be that he was particularly influenced by Hume in matters of moral philosophy, even if (I’m suggesting) ‘Freedom and Resentment’ in particular is instructively read as in important respects an extension of Strawson’s other writings.

engage in practices of responsibility—and that we, philosophers, can come to a better understanding of these—“*without worrying anymore about free will or the threat of determinism*” (D. Shoemaker, 2020, p. 212). The basis of this sense is not any one particular argument. It stems rather from the *approach* or *vision* or *attitude* which Strawson’s paper emanates.

This is *Strawson’s Promise*. This is what the paper evokes in some, which provokes others, and what is probably the most significant reason why the paper has been so influential and important. While mindful of the risk that a too stilted presentation may itself destroy the sense of promise, we may, however imperfectly, represent Strawson’s Promise a bit more concretely as involving the following two claims:

- (I) That determinism is irrelevant for responsibility.
- (II) That we may move beyond the traditional worries about free will.

This thesis sets itself to investigate how we may make good on Strawson’s Promise. A few previous readings, insofar as they themselves may be said to attempt this, will be assessed. The most promising route for making good on Strawson’s Promise, I’ll argue, is by a plausible understanding of Strawson’s Inescapability Claim—that is, the claim that we are naturally and inescapably committed to the concept of responsibility.

Strawson’s fundamental position regarding responsibility is, as the title of this thesis reads, *Minimal Optimism*. The optimism concerns our understanding of ourselves as responsible beings within a naturalistic understanding of the world, which is based on the indispensability of so understanding ourselves. The minimality consists in the fact that what our particular conception is like is underdetermined. What is inescapable is only that we have *some* conception of ourselves as responsible beings; what conception we have, or are to have, is in this respect a secondary question. Minimal Optimism is based on a recognition of the fact that any human society would need some conception of responsibility simply in virtue of very basic facts about beings like us.

There are a few different senses in which this suggests that determinism is fundamentally irrelevant for responsibility. For one, the very basic facts about human beings and human society in virtue of which the concept of responsibility is practically necessary hold irrespectively of the truth of determinism. Furthermore, the concept of responsibility that follows simply from these basic needs and situations is plausibly a rather metaphysically undemanding conception; we have no reason to think it would have to be in

conflict with determinism. Recognising, and coming to accept that, we are naturally and inescapably committed to the concept of responsibility, the question *whether* we are responsible is not our question anymore. The interesting question is, rather, what it means to be responsible. In this sense, Minimal Optimism brings us beyond the traditional debate.

III. The Optimist, the Pessimist, the Sceptic, and Strawson

This section briefly introduces how Strawson positions himself in relation to the other parties of the free will debate.

With the first two sentences of ‘Freedom and Resentment’, Strawson distinguishes between those who admit that they do not know what the thesis of determinism is and those who say, or imply, that they do. Strawson is of the first group. Of the second group, we get a further division of three subcategories in Strawson’s idiosyncratic terms: ‘the optimist’, ‘the pessimist’, and ‘the genuine moral sceptic’. What the optimist and the pessimist are respectively optimistic and pessimistic about is whether we are responsible agents if determinism is true.

The optimist is a *compatibilist*—someone who holds that our status as responsible beings is compatible with the truth of determinism. With the next section, ‘the optimist’ will be an *instrumentalist* or *consequentialist compatibilist*: someone who presents our practices of responsibility as essentially a behaviour regulation system and finds in this account the justification for holding people responsible. As most generally characterised, however, the optimist holds that the relevant concepts and practices “in no way lose their *raison d’être* if the thesis of determinism is true” (Strawson, 2008a, p. 1). Though the paper professes to be “a move towards reconciliation”, Strawson is a kind of optimist in these most general terms. Sufficiently—and that means *radically*—modified, the optimist is right, according to Strawson (2008a, p. 27). That does not necessarily or at least straightforwardly mean that Strawson is a compatibilist, as this view is standardly understood. It certainly does not mean that he is an instrumentalist or consequentialist compatibilist.

The pessimist is an *incompatibilist*—someone who holds that our status as responsible beings is incompatible with the truth of determinism. Under that heading, we find both *libertarianism*—the view that responsibility requires a kind of freedom incompatible with determinism, that determinism is false, and that we have this kind of freedom (at least sometimes)—and *hard*

determinism—the view that responsibility requires a kind of freedom incompatible with determinism, that determinism is true, and thus that we do not have this kind of freedom. A certain evaluative assumption on Strawson's part is evident already at this point: It is assumed that it is a good thing, that it is something that we cherish, that we are responsible agents. While libertarians are apt to think this too, hard determinists do not necessarily think that we are for the worse if we are not responsible agents, and might therefore wish to resist the label 'pessimist'.

The genuine moral sceptic is sceptical about responsibility whether determinism is true or not.¹⁶ Although the genuine moral sceptic makes only one more appearance in the paper, I think this is the central character of the plot. In more than one respect, Strawson is closer to the sceptic than he might seem. One respect is just that neither view is dependent on whether determinism is true or not. The sought-for reconciliation is between the optimist and the pessimist; furthest away from Strawson's synthesis is not the pessimist, then, but the genuine moral sceptic. But the genuine moral sceptic also sees something crucial, something Strawson also recognises: both the optimist's account and the pessimist's account of the relevant concepts and practices and their basis are fundamentally mistaken ("inadequate" or "inane" (2008a, p. 25)). Both the optimist and the pessimist seek a justification of our conceptual scheme and general framework of reactive attitudes as a whole—and this just can't succeed; the sceptic and Strawson are in fundamental agreement on this point. The difference is that the sceptic thinks that such a justification is necessary, while Strawson does not. This is the fundamental respect in which Strawson turns against all other parties of the debate. We'll consider this issue in some detail in chapter 2.

In the second section of 'Freedom and Resentment', Strawson briefly expands on the disagreement between the optimist and the pessimist. We are

16 McKenna & Pereboom (2016, p. 126) say that the sceptic, on Strawson's presentation, is a hard incompatibilist (i.e., someone that holds that free will is incompatible with determinism, that either determinism is true or a form of indeterminism is true which is also incompatible with free will, and therefore one has free will regardless of whether determinism is true or not (ibid. p. 32)). However, it is not entirely clear that the claim that "the notions of moral guilt, of blame, of moral responsibility are inherently confused" whether or not determinism is true commits Strawson's sceptic to any compatibility thesis. That they are "inherently confused" may rather echo Nietzsche's view that "causa sui is the best self-contradiction hitherto imagined, a kind of logical rape [... a] piece of nonsense" (2014, § 21). If we count 'the genuine moral sceptic' to those who say they know what the thesis of determinism is, then we should perhaps say that he is a hard incompatibilist. However, as 'the genuine moral sceptic' makes his second appearance, this aspect of the presentation is not salient, but the claim of nonsense is.

presented with a negative characterisation of the freedom necessary to be responsible—roughly, the absence of constraint, in the terms of classic compatibilism. We are also given a positive characterisation of the necessary freedom, in terms of the agent acting for reasons, not mere rationalisations. From this section and onwards, until the very last paragraph, ‘the optimist’ takes the more specific form suggested earlier; in this more specific form, the optimist is an instrumentalist or consequentialist compatibilist. The optimist holds that “the general reason” why our practices are as they are is that this is the way they need to be if they are to be generally efficacious in regulating behaviour in desirable ways. In reaching for this “general reason”, a familiar—at the time, rather prevalent—compatibilism, in Strawson’s terms, “goes beyond the facts as we know them”. That they feel this need to supply a justification is important; it is this felt need that Strawson seeks to free us from.

It is sometimes assumed that, contrary to the instrumentalist who thinks that “the efficacy of these practices in regulating behaviour in socially desirable ways” provides an adequate basis for our responsibility practices, Strawson’s central point is that this is “not a sufficient basis” and, more importantly, “not even the right *sort* of basis” for these practices (2008a, p. 4).¹⁷ While it is not mistaken that Strawson indeed thinks this, that is not the fundamental point of disagreement.

Rather, as Strawson hints at already in the second section of the paper, the issue concerns the need to identify “the general reason” why we are justified in holding each other responsible for what we do. The pessimist’s weakness for this kind of story expresses itself in their proneness “to alternate between the very familiar and the very unfamiliar”, between saying that justified blame requires that the person deserves the response and then that this requires “a genuinely free identification of the will with the act” (2008a, p. 3). Both the optimist and the pessimist feel the need to go beyond the facts as we know them. That is the fundamental misconception of the role of these concepts and practices in our lives which Strawson seeks to free us from.

What we need to do, Strawson argues, is to recall what it is actually like to be involved in interpersonal relationships with others and be mindful of “the kind of importance we attach to the attitudes and intentions towards us of those who stand in these relationships to us, and of the kind of *reactive* attitudes and

17 Darwall (2006) takes this to be “Strawson’s Point”. (Following Darwall, Shoemaker (2022, p. 309) suggests that Strawson’s eschewal of ‘external justification’ is a declaration of some considerations, such as the efficacy of holding others responsible in regulating their behaviour, as the wrong kind of reasons.) However, it is evident that Strawson is not speaking in his own voice here, but is (only) giving voice to the pessimist’s concern (cf., Miller, 2014a).

feelings to which we ourselves are prone” (2008a, p. 6). It is to the role of these concepts and practices in our lives that we must direct our attention. This is why Strawson introduces the philosophical community to *the reactive attitudes*—for example, “gratitude, resentment, forgiveness, love, and hurt feeling” (2008a, p. 5). Though what the reactive attitudes are is not very precisely defined by Strawson himself, it might be that the issue is best left that way. It might be that the terms are already by Strawson *too* defined.¹⁸

It is of note both that what Strawson seeks to remind us of are *commonplaces* stated in rather general terms and that he seeks to *remind* us of these. As he puts it in the introduction of *Individuals*, “They are the commonplaces of the least refined thinking; and are yet the indispensable core of the conceptual equipment of the most sophisticated human beings” (1959, p. 10). The thing is that we are all too prone to forget about them “when we are engaged in philosophy” (2008a, p. 7). What is needed is a more concrete and realistic understanding of the actual lives of our concepts.

III. Overview of Chapters

There are two structurally even parts to this thesis. The first part, consisting of chapter 2, 3, and 4, is negative. The second part, consisting of chapter 5, 6, and 7, is positive. They are respectively negative and positive as concerns Strawson’s Promise: the first part assesses previous attempts to make good on Strawson’s Promise, and shows that they are, in some or other respect, unpromising; the second part tries to explicate central features of Strawson’s approach as to show them to be more plausible than generally thought, hoping to thereby find a way to make good on Strawson’s Promise. There are certainly respects in which this presentation is not quite accurate. Neither part is wholly negative or positive. All chapters attempt to further our understanding of Strawson’s approach. They typically seek to problematise, qualify, or surpass some existent reading. But Strawson’s Promise is the theme of this thesis, and while not all chapters are directly concerned with it, they have been given the place they have in this thesis because they serve as parts in seeing if and how we may make good on Strawson’s Promise.

¹⁸ It is presented as something like an imperfection that “to some extent my own [Strawson’s] descriptions of human attitudes have reflected local and temporary features of our own culture” (2008a, p. 26).

Chapter 2, which takes its title from the first chapter of Strawson's *Scepticism and Naturalism*, 'Scepticism, Naturalism, and Transcendental Arguments', discusses scepticism, naturalism, and transcendental arguments. One view—which probably is, or has been, more widespread than it may seem—is that Strawson develops *a transcendental argument* against the responsibility sceptic.¹⁹ Justin Coates (2017) has proposed an account of Strawson's argument in 'Freedom and Resentment' as of this sort. Pamela Hieronymi (2020) also proposes that a transcendental argument is central to Strawson's approach. Given that Strawson famously developed transcendental arguments against other kinds of scepticism, the thought is not the least far-fetched. It might be that a kind of transcendental argument shows that responsibility scepticism is irrational, unreasonable or senseless; it might be that such an argument is implicit in, or otherwise supports, Strawson's approach in 'Freedom and Resentment'. But, so I argue in chapter 2, the argument Coates proposes is in any case not Strawson's argument. It is a transcendental argument *of the wrong kind* and, more fundamentally, expressive of *the wrong kind of* general anti-sceptical approach. This is not how to make good on Strawson's Promise, at least not as *Strawson's* Promise.

This chapter also serves as an introduction to Strawson's anti-scepticism. An interpretative puzzle regarding this is, what we'll call, *Glock's Question*: What is the nature of the sceptic's mistake, according to Strawson? The issue stems from the fact that Strawson seems to characterise the sceptic's mistake in rather disparate ways, corresponding to different anti-sceptical strategies: the transcendental and the naturalistic. By answering Glock's Question, we see how the two strands in Strawson's thought relate. This works as a foil to elucidate Strawson's method of transcendental argument, which informs the response to Coates. It also occasions an initial presentation of what Strawson's naturalism is; in particular, what its anti-sceptical potential is. It is with Strawson's naturalistic approach that we will be concerned in the latter part of this thesis.

Chapter 3, 'The Incoherence Puzzle', considers Pamela Hieronymi's novel reading of 'Freedom and Resentment'. She justifies this novel and self-consciously controversial reading by the injunction that we follow the text and take what Strawson actually says seriously; indeed, that we take it at face value. According to Hieronymi, the most straightforward understanding of what she calls "the central argument" of 'Freedom and Resentment' has been dismissed

19 In her obituary of Strawson, Jane O'Grady (2006) states, without further argument (it is an obituary after all), that 'Freedom and Resentment' consists of "a Kant-type transcendental argument" against the practical relevance of determinism.

only because it seems too implausible to be true. Hieronymi's account is an attempt to show that it is not, actually, that implausible; indeed, that it actually makes for "a powerful argument" (Hieronymi, 2020, p. 16).

We will not consider her account directly. That is, whether or not the view Hieronymi extracts from 'Freedom and Resentment' is implausible or powerful will not be our concern. We will consider the reading *as a reading*. We will take seriously Hieronymi's injunction to follow the text. Hieronymi claims that none of the existent readings can make sense of the fact that Strawson accuses his opponent of incoherence or self-contradiction. If we do take this interpretative puzzle seriously, however, then we encounter further and further interpretative puzzles. Why does he accuse the sceptic of incoherence *only once*? We should expect that he would do so at least twice. And is not Strawson, on Hieronymi's reading, contradicting himself? Rather than easing our puzzlement, the reading Hieronymi proposes, once its implications are clear, rather seems to increase it. As a reading of Strawson's approach, and of 'Freedom and Resentment' in particular, Hieronymi's reading is therefore considered unpromising.

Chapter 4, 'The Reversal Move', considers what the Reversal Move is, what it has been and is supposed to be, and then discusses two prominent accounts of it. The Reversal Move, broadly speaking, is said to be that Strawson reversed the order of explanation: rather than, as on 'the traditional approach', thinking that the *concept* of responsibility or what it means to *be* responsible determines or fixes our *practices* of responsibility or when to *hold* someone responsible, the relation is the other way around; it is the *practice* of responsibility, us *holding* people responsible, that (somehow) determines or fixes the *concept* of responsibility or what it means to *be* responsible. While there is something to this thought—something like it does shape Strawson's approach in a fundamental way—there are reservations to be had about the existent attempts to articulate this idea as a distinct thesis. What kind of explanation is at issue? How, exactly, does this contrast to the traditional approach? It is, however, with respect to the Reversal Move that the ambition to make good on Strawson's Promise is most pronounced. We need to see how it fares in this respect.

We first consider the history of the Reversal Move—i.e., the very recent history of it first being attributed to Strawson and what, then, it was taken to be. The hope is that this gives us a clue of what it was, and is, supposed to be, as well as what any thesis needs to be, first of all, in order to be a Reversal Move and, furthermore, in order to deliver on Strawson's Promise. We consider David Shoemaker's (2017, 2022) *Fitting Response-Dependence Thesis* and Benjamin De Mesel's (2022a) (and Sybren Heyndel's (2019))

Meaning-Based Account. The conclusion is that, unless these approaches involve a *naturalistic explanation* of our practices of responsibility that give us reason to think that determinism is not a relevant consideration for responsibility, which none of them do, then they will not make good on Strawson's Promise.

Chapter 5, 'The Inescapability Claim', explores the possibility that the least celebrated aspect of Strawson's response to the responsibility sceptic might in fact be the most promising way of moving beyond the traditional debate. A central tenet of Strawson's approach to the issue of free will and responsibility is *the Inescapability Claim*: the claim that we are naturally and inescapably committed to the concept, practice, and attitudes of responsibility. However, even by those otherwise sympathetic to Strawson's approach, the Inescapability Claim is generally considered not very plausible or convincing. The Standard Reading of the Inescapability Claim treats it as a psychological thesis which holds that having reactive attitudes is psychologically necessary for human beings in virtue of our psychological constitution. The Standard Reading, however, is mistaken. This chapter seeks to replace it.

By refocusing the Inescapability Claim, treating it as one among other affirmations of framework commitments that the descriptive metaphysician seeks to lay bare, we recognise *what* is inescapable and in what *sense*, or *senses*, that it is inescapable. The central argument of this chapter is that Strawson's view is a *Minimal Optimism*. On this reading, what is said to be inescapable is the concept of responsibility, not any particular conception of responsibility. In other words, what is inescapable is that we have *some* conception of responsibility; it is not therefore the case that any particular conception of responsibility is inescapable. The concept is inescapable in two different senses: it is practically necessary for the existence of human society, given the needs and situations of human beings; it is conceptually necessary if we are to have anything that we can make intelligible to ourselves as a system of human relationships. The practical necessity is in a sense more basic, at least with respect to responsibility.

Chapter 6, 'The Unanswered Question', considers a question that Strawson raised but never answered. Why do we have the concept of responsibility? This is the unanswered question of 'Freedom and Resentment'. By Strawson's own lights, we need to ask "how the nature of our thinking is rooted in the nature of the world and in our own natures" (Strawson, 2011b, p. 36). The kind of answer that Strawson should have given, I argue, is a *pragmatic genealogy*. Presenting the method of pragmatic genealogy, we consider in some detail the practice of blame as well as, in less detail, some further features of the concept of responsibility. The pragmatic genealogy shows that, for beings like us, the

concept of responsibility is inescapable, fundamentally, in the sense that it is indispensable. The pragmatic genealogy not only supports the Inescapability Claim, but it also renders intelligible why the concept of responsibility is practically necessary for beings like us already on the assumption only of very basic needs and situations of human beings and of human society. It thereby serves to *vindicate* us in living by the concept of responsibility.

Our pragmatic genealogy of the practice of blame furthermore reveals that, already in its highly generic form, it is a *self-effacingly functional* practice: it is functional, but only if the practice is guided by motives and reasons that are independent of the functionality of the practice in any given case. Given that the generic practice of blame is practically necessary, it is furthermore practically necessary that we have a conception of responsibility that is not purely instrumental; any actual conception of responsibility must be importantly backward-looking. The pragmatic genealogy not only allows us to get a clearer grasp on what the concept of responsibility already in its non-historical form must be like. With its State-of-Nature model, it also allows us to make naturalistic sense of the idea that the concept of responsibility is *non-historical*. Beyond supporting the Inescapability Claim, the pragmatic genealogy also serves to make good on Strawson's Promise in the sense that it shows that the concept of responsibility is practically necessary for beings like us already in virtue of needs that we have whether determinism is true or not and that the concept of responsibility that is necessary simply in virtue of these needs is plausibly itself metaphysically undemanding and does not require the falsity of determinism.

Chapter 7, 'Minimal Optimism & Strawson's Optimism', ventures beyond Strawson's Minimal Optimism. For one thing, it explores a critical edge to the Minimal Optimist view. The distinction between the concept of responsibility and any conception of responsibility thwarts the tendency to privilege our own conception as *the* conception. We thus avert mistaking scepticism about our conception of responsibility for scepticism about the concept of responsibility. While allowing for scepticism about our conception of responsibility, Minimal Optimism thwarts the tendency of such scepticism about responsibility to become *nihilism* about responsibility. This highlights a way in which Minimal Optimism changes the terms of debate.

A further respect in which we will go beyond Strawson's Minimal Optimism is that we will consider his optimism about our particular conception of responsibility. Admittedly, Strawson does seem to be not *only* a Minimal Optimist; he seems, at times, to endorse a more substantial optimism about responsibility. We will look specifically to the exchange between Jonathan Bennet (1980) and Strawson (1980). A few different readings of Strawson's

response to Bennett are entertained. One intriguing reading is that Strawson takes the idea behind his Minimal Optimism to be *analytically useful* in grounding an optimism about our particular conception, awarding us with a more concrete and realistic understanding of our ordinary notion of responsibility. At the end of this chapter, we also look at some reasons to think that Strawson's approach is more critical and more revisionary than he himself lets on.

Chapter 8 concludes this thesis. After a very brief summary, in light of what has come before, a concise articulation of the basic sense in which Strawson's Minimal Optimism might be said to make good on Strawson's Promise is given. What Strawson's approach effects is a change of terms. The question is not *whether* we are responsible. We cannot but be. Rather, the question is what it *means* to be responsible. Thanks to Strawson, *this* is our question.

2. Scepticism, Naturalism, and Transcendental Arguments

I. Introduction

Hans-Johann Glock (2022, p. 448), in his authoritative exposition of Strawson's general philosophical approach, poses the following question for any interpretation of Strawson interested in his anti-scepticism:

Glock's Question: What is the nature of the sceptic's mistake, according to Strawson?²⁰

The pertinence of Glock's Question derives from the fact that Strawson seems to characterise the mistake of the sceptic in two rather disparate ways. Sometimes, the accusation seems to be that scepticism is "meaningless" or "senseless" (e.g., Strawson, 1998c, p. 370). Other times, however, the accusation seems to be that scepticism is "idle, unreal, a pretense", "useless", or "in vain" (e.g., Strawson, 2008c, pp. 15, 16, 22). The disparity of these characterisations grows starker considering that Strawson himself makes a point of contrasting them:

It is to be remembered that the point has been, not to offer a rational justification of the belief in external objects or other minds, or of the practice of induction, but to represent sceptical arguments and rational counter-arguments as equally idle—*not senseless, but idle*—since what we have here are original, natural,

²⁰ I've reformulated the question slightly. Glock (2022, p. 448) asks about "the nature of this self-stultification"—i.e., of the self-stultification that the sceptic is charged with. However, it is only transcendental arguments that charge the sceptic with self-stultification; "the way of Naturalism" (2008c, p. 8), as Strawson calls it, is not to level this charge but to see the sceptical challenge as "idle, unreal, a pretense" (2008c, p. 15). The original formulation of the question does not really allow for the contrast that Glock intends for it to highlight and which he rightly calls our attention to.

inescapable commitments which we neither choose nor could give up. (Strawson, 2008c, p. 21f., emphasis added)

The plot thickens further considering that, elsewhere, the contrast seems to work in just the opposite direction:

It is not merely a matter of dismissing the demand for a justification of one's belief in a proposition on the ground that one can't help believing it. That would be weak indeed. The position is, rather, that the demand for justification is *really senseless*. (Strawson, 1998c, p. 370, emphasis added)

Strawson seemingly oscillates between two rather different characterisations of the sceptic's mistake: roughly, between the claim that scepticism fails because it is "idle" and the claim that it fails because it is "senseless". Each gloss is associated with a different strand in Strawson's philosophical methodology, both represented in this chapter's title: naturalism and transcendental arguments, respectively.²¹

Against this background, Glock elaborates on the interpretative question as follows:

Is [the sceptic's mistake] of a merely psychological kind, as a naturalist gloss would have it? Or does it have an argumentative, rational side, as the transcendental strategy demands? Is a sceptic merely paralyzed in his thinking and acting? Or is there also an intellectual flaw in his position? (Glock, 2022, p. 448)

How are we to make sense of this set of ostensibly conflicting and contrasting claims? It is the aim of this chapter to answer this question.

Regarding the substantive merits of either approach, transcendental arguments may indeed appear superior—on that strategy, we at least "silence the sceptic [...] by means of argument" (Glock, 2022, p. 449).²² Given the widespread dissatisfaction with "the naturalist gloss" on the sceptic's mistake, our chances of making good on Strawson's Promise may seem more hopeful on the transcendental strategy. Just this thinking has encouraged some to propose, *as a matter of interpretation*, that Strawson's argument against the

²¹ Adopted from the first chapter in Strawson's *Scepticism and Naturalism*. As Heyndels (2020, p. 100, n. 16) observes, the preface of the book states that the book comprises five lectures (even if the book is structured into four chapters), but there was a sixth lecture, "Causation and Explanation", which appears in *Analys et Métaphysique* (1985)/*Analysis and Metaphysics* (1992a).

²² This is Glock's view (see also, 2012, pp. 413-417; 2017, pp. 224-226).

responsibility sceptic is, or rests on, a transcendental argument. Below (section III), we will consider the most explicit account of this kind, as formulated by Justin Coates (2017).

In addition to the exegetical issue of Glock's Question and how to make sense of the seemingly conflicting claims it points to, there is a further question—not *of* interpretation but *for* interpretation—that we need to consider. We are primarily concerned with scepticism about responsibility and Strawson's approach to this kind of scepticism. There are many other kinds of scepticism (about the external world, other minds, or inductive reasoning, for example) and Strawson discusses these other kinds of scepticism as well. How different are these "kinds of" scepticism? Is scepticism always roughly of the same form, only *about* different issues of philosophical reflection; alternatively, always stemming from the same source, only manifesting itself in different forms? Could there be a universal anti-sceptical method that is (in principle) invariably applicable, such that we may seek one method to overturn all philosophical scepticism(s)? Or is, for example, scepticism about the external world and scepticism about responsibility importantly different issues, such that the pursuit of a sweeping response to scepticism is only a pipedream of some all too intellectually complacent philosophers? For the project of rational reconstruction, it is imperative that we ask: Can we legitimately draw on Strawson's response to, say, scepticism about the external world in order to further explicate his response to responsibility scepticism?

The argument of this chapter is that Strawson's approach to scepticism is both *universal* and *discriminate*. That is, he *both* treats all scepticism in the same way *and* responds differentially to different sceptical challenges. The paradoxical appearance of this claim is only superficial; the universality and the discriminating nature of Strawson's approach to scepticism may be accommodated as compatible and indeed complementary parts of Strawson's anti-scepticism. The universal anti-sceptical aspect of Strawson's approach is encapsulated in his brand of naturalism. The discriminating aspect of his anti-scepticism comes to the fore when we consider his employment of transcendental arguments. That there is this combination of a universal and a discriminating nature to Strawson's anti-scepticism is not duly recognised in the literature; the assumption is most often that Strawson's treatment of scepticism is (only) of the sweeping (one-response-to-all-challenges) variety. Therefore, the need to dispel what seems paradoxical in the here argued for claim typically does not even arise.

Still, it is recognised that we need to answer Glock's Question; any reading of Strawson must grapple with the issue of somehow accommodating claims and arguments that either seem to be in conflict or are themselves claims of

their being a kind of conflict between different responses to scepticism. At least implicitly then, there need be an assumption about the compatibility of Strawson's different anti-sceptical stripes.

It is also generally held that Strawson not only contrasts naturalism with transcendental arguments but explicitly opts for the former *as an alternative* to the latter.²³ Observations like these have encouraged suggestions such that there is a "turn" in Strawson's thought (Stern, 2003); an inconsistency in his treatment of scepticism (Glock, 2022, p. 449); or an abandonment of transcendental arguments, or of a particular form of such arguments (Callanan, 2011, pp. 262, 273) and, especially, of their anti-sceptical potential. Once the relation between naturalism and transcendental arguments in Strawson's approach to scepticism is properly appreciated, however, we will see that these exegetical proposals are misguided.

A proper appreciation of Strawson's method of transcendental arguments and its relation to his naturalism, which this chapter aims to effect, will allow us to answer Glock's Question, make sense of the apparent conflict, and remove the misleading alternative interpretations.

II. Transcendental Arguments

In this section, we consider the ambitions and limitations of transcendental arguments. One appeal of transcendental arguments is that they hold anti-sceptical potential; especially, in the context of Strawson's philosophy, that they hold greater anti-sceptical potential than naturalism does: to show that the sceptic's challenge is *senseless* or *incoherent* is to show a serious conceptual, intellectual, or rational flaw in their argument; this may indeed seem a more serious charge than merely replying to the sceptic that their doubt is "idle, unreal, a pretense". This, in any case, appears to be the general opinion.²⁴ It is, as we'll see, also Strawson's opinion. Nevertheless, Strawson recurrently opts for 'a naturalist response' (and we will soon consider in more detail how to

²³ This is the passage encouraging that reading: "Is there any other way with scepticism which is not a variant on those I have referred to, i.e. is neither an attempt directly to refute it by rational argument drawing on commonsense or theological or quasi-scientific considerations nor an attempt indirectly to refute it by showing that it is in some way unintelligible or self-defeating? I think there is another way. [...] I shall call it the way of Naturalism" (Strawson, 2008c, p. 8).

²⁴ Consider, for example, Glock (2022), Coates (2017), and Hieronymi (2020).

properly understand that). Despite their greater potential, then, why *not* transcendental arguments?

A transcendental argument is characterised by a certain kind of claim—the *transcendental claim*—that *p* is a necessary precondition for the possibility of *q*. Transcendental arguments are typically developed against scepticism of some sort: the transcendental claim as against *p*-scepticism (scepticism about *p*) is meant to show that something that the sceptic cannot sensibly deny (viz. *q*) presupposes *p* (i.e., what the sceptic doubts). The hook of a transcendental argument is that the sceptic is said to be somehow antecedently committed to *q*, such that the sceptic's denial of *p*, which thereby implies a denial of *q*, is somehow mistaken. The qualificatory 'somehow's are necessary because depending on the *kind* of transcendental argument, the antecedent commitment varies (if it is having meaningful discourse in general or of some kind, or is that of having a particular ability or some subjective experience, for example) and so does the mistake (being, for example, of veracity or rationality).²⁵

To properly discern the relevant kinds of transcendental arguments, a distinction between anti-sceptical arguments in general is useful: An anti-sceptical argument may be either *direct* or *indirect*.

A *direct* anti-sceptical argument typically seeks to show that the sceptic's claim is *false*. While the sceptic denies *p*, a direct anti-sceptical argument seeks to show that *p* is or must be the case—directly refuting the sceptic's claim. G. E. Moore's (1939) "Proof", from knowledge of his hands (and that they are external objects) to knowledge of the external world, on a popular understanding, is taken to be a typical example of an anti-sceptical argument that seeks to *directly* refute the external-world sceptic's claim.

An *indirect* anti-sceptical argument, on the other hand, does not seek to directly refute the sceptic's claim, but nevertheless aims to undermine the sceptical challenge in some other, more indirect way. For example, it may seek to undermine scepticism by exposing and rejecting *the underlying allure* of the scepticism, or by showing that the sceptical *challenge* itself is somehow confused or cannot be coherently stated (without thereby directly refuting the sceptical conclusion).

²⁵ A further question is, in what sense is *p* presupposed—i.e., in what sense is *p* supposed to be necessary for the possibility of *q*? It is sometimes glossed as an *a priori*, *non-empirical*, or *conceptual* necessity. In any case, on a transcendental argument, *p* is not taken to be merely a condition on the possibility of *q* in a *causal* sense. For our purposes, we may (following Anil Gomes (2017, p. 135)) settle for a characterisation of the nature of the necessity as *constitutive*: *p* is a precondition on the possibility of *q* in the sense that *p* is *constitutive* of *q*. This is ecumenical enough without being vacuous.

The paradigmatic kind of transcendental argument belongs to the family of direct anti-sceptical arguments—this is the *deductive* transcendental argument.²⁶ A deductive transcendental argument seeks to deduce the falsity of scepticism—i.e., deduce that *p* from the fact that *q* and a transcendental claim (that *q* necessarily presupposes *p*). Beside the transcendental claim, the further premise is an affirmation of *q*, which may be a (non-controversial) belief about the world or claim about us (for example, that we have a particular kind of experience or ability); typically, *q* is the claim that we have some particular experience that—importantly—not even the sceptic denies we have. Schematically, deductive transcendental arguments take the following form:

- P1: We have some belief/ability/experience (*q*)
- P2: It is a necessary precondition for having that belief/experience/ability (*q*) that *p* is the case [the transcendental claim]
- C: Therefore, *p*.

If sound, this makes for a proof against scepticism: the sceptic's claim must be plain false or at best irrational (e.g., epistemically unfounded). This is a key characteristic of a deductive transcendental argument: from something the

²⁶ Another categorization—although, more specifically, of *kinds of transcendental arguments* (rather than anti-sceptical strategies in general)—is that between *truth-directed* or *world-directed* arguments on the one hand and *belief-directed* or *self-direct* arguments on the other (cf., Cassam, 1997, p. 78f.; 1999; Gomes, 2017, p. 133; Heyndels, 2020, p. 126; Pereboom, 2016, 2022). I will use the direct/indirect distinction with respect to anti-sceptical strategies in general and the deductive/elenctic with respect to transcendental arguments, following Callanan (2011) and Glock (2003b, 2016, 2022). Strawson (2008c, p. 8) himself also employs the direct/indirect distinction with respect to kinds of anti-sceptical strategies. Admittedly, though, the phrasing in terms of “deductive” is somewhat misleading. The difference between deductive and elenctic is essentially in whether the argument is foundationalist-friendly or not; if it awards a direct or an indirect refutation of scepticism. The difference is not essentially in the logical structure of the argument. However, the anti-scepticism, as we will see, might actually not be essential, at least for elenctic transcendental strategy, so this contrast also halts. Most probably, the label ‘deductive transcendental argument’ is explained by the arguments’ inspiration: Kant’s Transcendental Deduction. (As others have noted, Kant does not himself use the term ‘transcendental argument’ in its contemporary sense. He speaks of ‘transcendental deductions’ and similar which means about the same as ‘transcendental argument’ today (Stern, 2019).) The general method, however, traces back, at least in some form, already to Aristotle’s argument against the sceptic about the principle of non-contradiction (*Metaphysics* 1005b35-1006a28). But Kant is still the prime example; the increased popularity of transcendental arguments among analytic philosophers in the 20th century was synchronous with the rise of analytic Kantianism. (On the rise of analytic Kantianism and the revival of metaphysics in the 20th century, particularly in relation to Strawson, see Hacker (2001) and Glock (2003b, 2012).)

sceptic does (or, at least supposedly, needs to) accept (viz. *q*), a connection of presupposition can be established such as to deduce the incorrectness (falsity or irrationality) of the sceptic's claim. This amounts to a direct refutation of scepticism about *p*.

A salient line of response for the sceptic is to question the transcendental claim, whether *q* really presupposes *p*. Beside this common response, another response characteristic of *deductive* transcendental arguments in particular is that the sceptic double down: if *q* really presupposes *p*, then we need to also reject *q*. ('One man's *modus ponens* is another man's *modus tollens*.') Even more characteristic of the dialectic of deductive transcendental arguments is that the proponent of the transcendental argument then returns the favour. Evidently, this iteration of the argument goes, the sceptic is irrational: rather than giving up their initial scepticism, the sceptic is willing to extend their doubt (irrationally enough) to even the most plausible, uncontested assumption (i.e., *q*). By the lights of a deductive transcendental argument, scepticism is false or, at best, irrational.

Deductive transcendental arguments have received substantial criticism over the decades. The *locus classicus* of this criticism is Barry Stroud's (1968) "Transcendental Arguments".²⁷ Most notably, Stroud's criticism takes as its target Strawson's (1959) use of transcendental arguments against scepticism about the external world and about other minds.²⁸ Stroud's main objection is that transcendental arguments "would have to show that [...] a statement about the way things are, follows from [...] a statement about how we think of the world, or what makes sense to us" (1968, p. 246). According to Stroud, "the most that could be proved by a consideration of the necessary conditions of language is that, for example, we must *believe* that there are material objects and other minds if we are to be able to speak meaningfully at all" (1968, p. 256). Since a transcendental argument can only show that since (or if) we *think* in a certain way, have some *experience*, or *believe* something (i.e., *q*), we thereby need to *assume* that *p* (the reality of the external world, for example), a transcendental argument leaves it open for the sceptic to "very plausibly

²⁷ Stroud's paper is regularly credited with the "disillusionment" of the promise of transcendental arguments (e.g., Callanan, 2011, p. 263; Stern, 2019, sec. 3). For a largely analogous critique, though of (an interpretation of) Wittgenstein's private language argument, see Thomson (1964). Notably, Stroud (1968, p. 255, n.12), in his paper, refers to Thomson; for discussion, see Rorty (1971). Glock (2003b, p. 38) observes that Stroud's criticism was in fact anticipated by C. D. Broad (1978) in his lectures on Kant in 1951, but take it that Stroud could not have known of these when he formulated his own critique.

²⁸ Stroud's criticism was not exclusively reserved for Strawson: it also targeted Shoemaker (1963).

insists that it is enough [...] if we *believe* that [*p*] is true, or that it looks for all the world as if it is, but that [*p*] needn't actually be true" (Stroud, 1968, p. 225, original emphasis). Stroud rightly insists on the gap between *what we think about the world* and *what the world is actually like*, and the criticism comes to the fact that transcendental arguments cannot, contrary to their pretensions, bridge that gap. They cannot show that what the sceptic denies is in fact the case (unless they, implausibly, rely on a simple form of verificationism (Stroud, 1968, p. 247)).

Stroud's criticism of transcendental arguments is premised on the assumptions, first, that the sceptic makes a claim about the way things *are* or us *knowing* about the way things are and, second, that the starting point of the transcendental argument needs to be a claim the sceptic accepts and is, typically, a claim about *how we think* (e.g., the meaningfulness of some discourse) or what we *believe*. The criticism follows because what the transcendental argument is out to establish is that the sceptic is mistaken in point of fact; that is, it follows on the condition that the transcendental argument under consideration is a direct anti-sceptical argument.

Considering Stroud's criticism, Strawson (2008c) straightforwardly concedes the point that "the most that could be proved by a consideration of the necessary conditions of language is that, for example, we must *believe* that there are material objects and other minds if we are to be able to speak meaningfully at all" (Stroud, 1968, p. 256). Stroud's point can simply be accepted by Strawson because the proper role of transcendental arguments, according to Strawson, is *not* to directly refute the sceptic's claim.²⁹ If transcendental arguments nevertheless can, as Stroud admits that they can, "demonstrate something about the use and interconnectedness of our concept" (Strawson, 2008c, p. 7), then that is, by Strawson's lights, enough.³⁰ Stroud misunderstands Strawson's intentions: Strawson's way with the sceptic is *indirect* rather than direct; he does not seek to deduce the falsity of the sceptic's

²⁹ It may be argued that Strawson *changed* his mind: that his early transcendental arguments *were* deductive but that he later *abandoned* this method. Among others, Stroud himself seems to think so (Stroud, 2002, fn. 14). I address this issue below.

³⁰ As he put it later on, "I am disposed to settle for that" (Strawson et al., 2008, p. 92); and, "even if some [transcendental arguments] lack the modal strength sometimes claimed for them, they may yet serve to bring our conceptual connections and dependencies of great interest and, as I think, importance" (Strawson, 1995, p. 415). In considering whether it might be thought that "Strawson's descriptive metaphysics yields transcendental arguments that prove the existence of the external world or other minds", Hacker takes this idea to be plain mistaken, adding that it would indeed "be absurd to argue from conceptual connections in thought to existential truths about the world, or, in Wittgensteinian idiom, from grammatical propositions to empirical ones" (2001, p. 363).

claim, but rather to elucidate the *incoherence* of the sceptic's challenge (Bakhurst, 2017, n. 16; Callanan, 2011, p. 269). The kind of transcendental arguments that Strawson develops are *elenctic* transcendental arguments.

Elenctic transcendental arguments purport to show the sceptical challenge itself to be *incoherent* or *self-refuting*. They are indirect in that they do not, by showing the incoherence of the *p*-sceptical challenge, thereby show the sceptic's claim to be false. They seek to show that the *argument* for the sceptic's claim is incoherent or self-refuting. Not only that the argument is unsound, but that it is senseless. This does not mean that the sceptic's claim is false. But it does mean that we have been given *no reason* to think the sceptic's claim is true, because they have not managed to make a meaningful argument.

The hook of such arguments is that the sceptic implicitly presupposes the validity of the prevailing conceptual scheme in raising the challenge while, per the sceptical conclusion, also claiming that this conceptual scheme is itself invalid. The sceptic employs concepts that make sense only if we assume that the conceptual connections that the sceptic explicitly denies nevertheless hold. The sceptical challenge could not be stated unless it were unfounded.

While deductive transcendental arguments purport to show that the sceptic's claim, which is a claim about how the world is (or isn't), is false, an *elenctic* transcendental demonstration shows not what the world is like but only how we must *think about* the world. This is nevertheless enough if the claim is that scepticism about our conceptual scheme (or, more precisely, a particular feature of our conceptual scheme) cannot be coherently proposed because the sceptical challenge itself relies on that very conceptual scheme. Because they are in this way indirect (or not world-directed), elenctic transcendental arguments are sometimes presented as more "modest" than deductive transcendental arguments.³¹

Some have thought that Strawson's "concession" to Stroud is substantive, representing a methodological shift.³² It might seem like Strawson (2008c)

³¹ On both Stern's (2019) and Bardon's (n.d.) overviews, as it is in Stroud's (1968) "Transcendental Arguments", the elenctic argument is the more "modest". I take them to label it thus because it does not have the ambition to be the foundationalist, direct refutation of scepticism that they implicitly suppose that we would want. (See also, Bardon, 2005; Stroud, 2002, p. 209f.).

³² Quassim Cassam (2008, p. ix) suggests that there was a shift (albeit, not unambiguously, since he also notes how congenial the indirect, elenctic transcendental strategy is to Strawson's descriptive metaphysics). Glock (2003b, p. 39), similarly, says that Strawson "subsequently accepted this criticism [of Stroud's]" and that he "concedes" Stroud's point, but also stresses the commonalities of elenctic transcendental arguments and Strawson's naturalism. Grundman and Misselhorn (2003, p. 206, emphasis added) attribute a similar change of mind, proposing that Strawson (as of *Scepticism and Naturalism*) "*no longer* believes that

himself admits as much in *Scepticism and Naturalism* (cf., Callanan, 2011, pp. 262, 273). But this is not right. The project of *Individuals* (where the particular arguments that Stroud discusses are to be found) is that of descriptive metaphysics: “to describe the actual structure of *our thought* about the world” (Strawson, 1959, p. 9, emphasis added). This programmatic statement gives us reasons to doubt that Stroud accurately characterises Strawson’s transcendental arguments as attempting “to show that [...] a statement about how things are, follows from [...] a statement about how we think of the world, or what makes sense to us” (1968, p. 246). Descriptive metaphysics—as philosophy in general, according to Strawson—is *an exercise in conceptual self-understanding* (Heyndels, 2020).³³ There is no inference from how we must think of the world to how the world is. The aim is simply to describe (the most fundamental features of) *how we think*—to describe, that is, the workings of our conceptual scheme.

To defeat any lingering suspicion about the claim that the kind of transcendental arguments Strawson employed were *elenctic*, as well as to further illustrate that method of argument, consider Strawson’s argument against the sceptic about material objects. (The following reconstruction follows John J. Callanan’s (2011, pp. 266-270).) In *Individuals*, Strawson argues that we are committed to the claim that particulars continue to exist while unobserved, since we at least sometimes successfully re-identify particulars despite discontinuity in our observation of the particular. The sceptic concedes the following premise:

We have the idea of a single spatio-temporal system of material things. (Strawson, 1959, p. 35)

What the sceptic doubts is our *ground* for re-identifying any particular within that system. According to the sceptic, “for any occasion where one has two successive observations at t_1 and t_2 , every second occasion of observation at t_2 could feasibly be an occasion of observation of an entirely new and different particular to the one previously observed” (Callanan, 2011, p. 267).

The sceptic is right that in any particular case of discontinuous observation, it is always possible that a particular observed at t_1 and a particular observed at t_2 are numerically distinct particulars; not the discontinuous observation of

transcendental arguments can refute the sceptic”. Also Walker (1989, p. 55). (cf., Callanan, 2011, p. 261). This perception is perhaps not wholly independent of that other methodological discontinuity claim, i.e. of a “naturalistic turn”.

³³ See for example Strawson (Pivčević, 1989; 1959, p. 9; 1966, p. 72; 1982, p. 50; 2008c, p. 17; 2011e, p. 225); also, Hacker (2001, p. 363f.) and Glock (2003b, 2012, 2022).

the same particular. However, according to the sceptic, that this is possible in *any* particular case generalises to a claim about *every* possible case. Thus, we *never* have grounds for re-identification. This is the sceptical conclusion. Is it coherent?

According to Strawson, it is not. On a general gloss, this is the sceptic's mistake:

the sceptic could not even raise his doubts unless he knew it to be unfounded; i.e. he could have no use for the concepts in terms of which he expresses his doubt unless he were able to know to be true at least some of the propositions belonging to the class of all members of which fall within the scope of the sceptical doubt.³⁴ (Strawson, 2008c, p. 7)

The "proposition belonging to the class of all members of which fall within the scope of the sceptical doubt" in the case of the sceptic about material objects is the notion of "a second observation after an interruption representing a continuation of a first observation" (Callanan, 2011, p. 267). If it was the case that we *never* have grounds for re-identification of particulars, as the sceptic claims, then we would "have the idea of a new, a different, spatial system for each new continuous stretch of observation" (Strawson, 1959, p. 35). And if that was the case,

there would be no question of *doubt* about the identity of an item in one system with an item in another. For such doubt makes sense only if the two [spatial] systems are not independent, if they are parts, in some way related, of a single system which includes them both. (Strawson, 1959, p. 35, original emphasis)

If the sceptical conclusion were correct, we would not have the doubt 'Is this observation of a particular at t_2 representing the continuation of an observation of a particular at t_1 ?' We would not have the doubt 'Is this the same particular or not?' But we *do* have such doubt. Most importantly, the sceptical challenge itself *presupposes* that we have such doubt; it could not otherwise even get started. Therefore, if the sceptic's conclusion were correct, the sceptical challenge could not have been meaningfully expressed: the sceptic's "doubts are unreal [...]" because they amount to the rejection of the whole conceptual

³⁴ More precisely, this is how Strawson characterizes the point of indirect transcendental arguments *in general*.

scheme within which alone such doubts make sense” (Strawson, 1959, p. 35).³⁵ This is not to say that there *are* material objects. But it is to say that the sceptical argument to the contrary is in this sense *incoherent*. In an apt metaphor, what an elenctic transcendental argument shows is that the sceptic is sawing off the branch on which he is himself sitting (cf., Glock, 2022, p. 448).

On a plausible reconstruction, Strawson developed *indirect, elenctic* transcendental arguments (already in *Individuals* (1959)), *pace* Stroud’s assumption, and thus, not only is Stroud’s criticism misplaced, but, furthermore, the idea of a *methodological shift* in Strawson’s approach from deductive to elenctic transcendental arguments is not plausible. *If* there is a transcendental argument in ‘Freedom and Resentment’, or if any such argument structures or informs Strawson’s approach to scepticism about responsibility (there or elsewhere), then we should expect it to be of the elenctic (so-called “modest”) kind.

III. A Transcendental Argument in ‘Freedom and Resentment’?

Given that Strawson developed transcendental arguments against forms of scepticism in *Individuals*, it is not a far-fetched thought that ‘Freedom and Resentment’, written shortly after the publication of *Individuals*,³⁶ might embody just that kind of argument. The idea that Strawson’s argument against the responsibility sceptic is a kind of transcendental argument has been explicated and defended by Justin Coates (2017).³⁷ Coates explicitly offers this

³⁵ Note the term “unreal” here in the concluding indictment of an anti-sceptical transcendental argument, it being a term that Strawson also uses for the naturalist retort to the sceptic’s doubts.

³⁶ By Strawson’s (1998a, p. 11) reminiscence, ‘Freedom and Resentment’ was “written about the same time” as ‘Social Morality and Individual Ideal’, which was written in 1960 (published 1961, but delivered as a lecture already in the fall of 1960). While Strawson was elected to the British Academy already in 1960 (see, “Annual Report, 1960-1” Proceedings of the British Academy” Annual Report, 1960-1,” 1961), ‘Freedom and Resentment’ was not delivered at the Academy until May 9th 1962; regarding when more exactly it was written, I have found no further information. I suspect a more exact dating will not be of importance for any argument of this thesis. I thank Nicola Zetti for pressing me to get clearer on Strawson’s reminiscence here.

³⁷ Beside Coates’s interpretation, Benjamin De Mesel (2018), in response to Coates, has suggested an alternative (as he sees it, corrected) transcendental argument as part of

reading in contrast to the more common “naturalist” reading (the Standard Reading) which takes the Inescapability Claim to be the central move.³⁸ Some of the professed appeal of Coates’s reading is the possibility of avoiding the (supposedly implausible) naturalist response to scepticism.

The difference between the Standard Reading and Coates’s new reading, most succinctly put, pivots on how to understand Strawson’s claim that it is “practically inconceivable” that the objective attitude completely dominates our lives such that there is no practice of responsibility, no use of the concept of responsibility, no occasion for reactive attitudes (*as* reactive attitudes). Regarding the complete abandonment of the practice, concept, and attitudes, Strawson says: “I am strongly inclined to think that it is, for us as we are, practically inconceivable” (2008a, p. 12). The Standard Reading understands Strawson as here taking this (responsibility sceptical) scenario to be a *psychologically impossibility*. Coates, however, understands Strawson as here taking the (responsibility sceptical) scenario to be, not psychologically impossible, but impossible because inconceivable “from the *practical point of view of agents engaged in ordinary interpersonal relationships*” (2017, p. 807, original emphasis). While the Standard Reading renders the complete abandonment of the practice of responsibility as an impossibility conditional on the nature of human psychology, Coates’s reading renders it impossible

Strawson’s strategy in ‘Freedom and Resentment’. John J. Callanan (2011, p. 276) has suggested that Strawson’s argument in ‘Freedom and Resentment’ is continuous with (is of “exactly the same” kind as) his argument against external-world scepticism in *Individuals* (1959)—which is a transcendental argument. Pamela Hieronymi (2020, p. 28) has suggested that Strawson, in ‘Social Morality and Individual Ideal’ (2008d), provides “the ingredients for a transcendental argument” which may refute the responsibility sceptic and which underlies his central argument in ‘Freedom and Resentment’. However, while noting that “the ingredients” are there and building from that, she does not clearly proceed to develop the argument *as* a transcendental argument (cf., Heyndels, 2022a, p. 6f.). Hieronymi’s repeated invocation of ‘conditions of possibility’ for responsibility does most plausibly stem from her reading Strawson as providing some transcendental argument, even if that is not made (sufficiently) explicit. We will consider Hieronymi’s reading in the next chapter. (For a reconstruction of Hieronymi’s argument as a transcendental argument, see Russell (2021, p. 760).) In her obituary of Strawson, as mentioned, Jane O’Grady (2006) also states that ‘Freedom and Resentment’ involves a transcendental argument.

³⁸ Coates accurately represents the Standard Reading as attributing to Strawson the claim that it is *psychologically impossible* that we abandon the reactive attitude or the practice and concept of responsibility. He takes the Standard Reading to furthermore suppose that the psychological necessity of the attitudes/practice/concept is meant to yield a justification of the practice. For this characterisation, Coates relies on Paul Russell’s (2017d) interpretation of Strawson’s approach.

conditional on the fact that we are practical agents engaged in ordinary interpersonal relationships.

We can schematically represent the transcendental argument Coates proposes as follows:

- P1: We engage in ordinary interpersonal relationships (*q*).
- P2: It is a necessary precondition for engaging in ordinary interpersonal relationships that we engage in the practice of responsibility.
- C: We legitimately engage in the practice of responsibility (*p*).³⁹

What Coates's transcendental argument aims to establish is the *legitimacy* of our engagement in the practice of responsibility. The relevant question is, as Kant called it, a "question of right": what is our *justification* for engaging in this practice (or using this concept or having these attitudes)?⁴⁰

The transcendental claim (P2) is that *p* is *constitutive* of *q*, such that we cannot be engaged in ordinary interpersonal relationships without being engaged in the practice of responsibility. Here, "cannot" expresses a requirement of practical rationality: the practice of responsibility "cannot be questioned from the perspective of agents who are actively engaged in [... ordinary interpersonal] relationships" (Coates, 2017, p. 807). That is, "our commitment to regarding one another via the reactive attitudes" (ibid.) stands as a *practical postulate* to our engagement in ordinary interpersonal relationships. From that practical perspective, according to Coates, "we are rationally committed to the legitimacy of [...] responsibility practices" (2017, p. 817). We may grant the claim that this kind of practical necessity is at least *prima facie* better suited to yield a *rational justification* of our responsibility practices than a 'natural' or psychological necessity is.

³⁹ On De Mesel's (2018) reconstruction of Coates's transcendental argument, it is an important fact that P1 reads, rather, something like "We are *legitimately* engaged in ordinary interpersonal relationships". This may be thought important if the argument is to be valid: if "legitimately" is not in the premises, how can it be in the conclusion? (This thought is what drives De Mesel's criticism of Coates's interpretation.) However, this thought underappreciates what P2 is meant to do: it is crucially the transcendental claim that is supposed to *generate* the justification, such that "legitimately" in P3 is not illicit simply because there is not a "legitimately" in P1. As this point brings out, it should be noted that the presentation of the transcendental argument offered here is too schematic in certain respects. But it is nevertheless not too schematic for our discussion.

⁴⁰ Kant *Critique of Pure Reason* (A 84 ff.); see Stroud (1968, p. 241).

Coates calls Strawson's argument, on his reading, a "modest transcendental argument". His presentation suggests two reasons for the label "modest".⁴¹ First, modesty is said to be a matter of scope: while some transcendental arguments (as Kant's and some Kantians') are more ambitious in that they claim to hold for *any possible agent*, "Strawson's argument, if successful, shows that moral responsibility is practically inescapable *only* for agents who are social *in a particular way*" (Coates, 2017, p. 817, original emphasis). Second, modesty is said to be a matter of anti-sceptical force: *p*-scepticism (in this case, scepticism about the practice of responsibility) fails "not because *p*-scepticism is somehow incoherent, but because *p* itself is an essential element of a larger practice whose justificatory status is not in question" (2017, p. 810). The transcendental argument is modest because it is, in who it addresses and in what sceptical challenge it can resist, *conditional* on an antecedent commitment; in this case, on our antecedent commitment to ordinary interpersonal relationships.

However, neither of these considerations warrants the label "modest" as this is used in the literature on transcendental arguments; i.e., "modest" as opposed to *direct* or *deductive* transcendental arguments (see e.g., Bardon, 2005, no date; Stern, 2019; Stroud, 1968, 2002).⁴² Any transcendental argument is conditional on *q*—this holds equally for deductive as for elenctic transcendental arguments. Thus, for all said so far, Coates's "modest transcendental argument" may be a *deductive* transcendental argument. And if it is, given the argument of the last section, the interpretation's claim to reconstructing *Strawson's argument* is dubious.

⁴¹ I say that his presentation *suggests* two reasons, rather than that he himself *provides* two reasons, because while I conceive of these as two different reasons (of scope and anti-sceptical force, respectively), it is not clear from the text that Coates would himself distinguish the two considerations, or for that matter distinguish them as such.

⁴² Sometimes, Coates's characterisation seems not just different, but to be the opposite one as that of the traditional literature on transcendental arguments: "The most ambitious form of transcendental argument" according to Coates, "will attempt to show that *p*-scepticism is *incoherent*" (2017, p. 809', emphasis added). But that is the goal of an *elenctic* transcendental argument, the kind of argument traditionally characterised as the "modest" form of transcendental argument. Strictly, though, elenctic transcendental arguments seek to show the sceptical *challenge* to be incoherent; scepticism is incoherent in the sense that *the reasons for* scepticism are incompatible with *the conclusion of* scepticism. Notably, at one point, this is how Coates characterises transcendental arguments *in general*: transcendental arguments in general are said to attempt to show that "the *p*-sceptic must presuppose some element of *p* in order to motivate her scepticism, [to show that ...] she has no legitimate grounds for her scepticism" (Coates, 2017, p. 809).

The way Coates's defends the argument tells of it *not* being an elenctic or otherwise indirect argument, but rather a *direct* anti-sceptical argument. As mentioned, there are roughly speaking two lines of response to a deductive transcendental argument: deny either P1 (questioning *q*) or P2 (questioning the transcendental claim). Whether engagement in ordinary interpersonal relationships indeed does presuppose engagement in the practice of responsibility is not obviously so and requires some argument.⁴³ The plausibility of any transcendental argument of course depends on how credible the demonstration of the transcendental claim can be made. Characteristic of a *deductive* transcendental argument, however, as noted, is the other sceptical response—that the *p*-sceptic double down: granted the transcendental claim, they simply follow the argument and resolutely deny not only *p* but also *q*. And even more characteristic of the dialectic of deductive transcendental arguments is that the proponent of the argument returns the favour by simply inflating their criticism of the sceptic: the sceptic must surely be irrational for they are willing to extend their doubt to the most plausible assumptions. To the point: This is just the dialectic Coates is involved in. Of the sceptic that is ready to meet his argument in this way, he asks:

is it really credible that our friendships and mature love relationships are *per se* illegitimate? Or is it instead more plausible that such relationships are valuable, and so worthy of the import we naturally attach to them? (2017, p. 818)

This dialectic suggests that Coates's "modest transcendental argument" is in fact a deductive transcendental argument. It is hence *the wrong kind of* transcendental argument—i.e., the kind that Strawson agrees (with Stroud and others) are not to be employed.

Now, whether or not the argument perfectly conforms to a deductive transcendental argument, it is in any case a *direct* anti-sceptical argument. On Coates's reading, "Strawson is providing a transcendental *justification* for our responsibility practices" (2017, p. 808, emphasis added). The argument aims to directly refute the sceptic; in Strawson's words, such an argument is an attempt at "supplying the reasoned rebuttal which the sceptic perversely

⁴³ See Coates (2017, p. 813ff.). Coates explicitly opposes Pereboom's (2001) and Sommers' (2012) scepticism about this claim (See Pereboom (2016, 2022) for a very brief retort directly to Coates.) For further discussion on this point, see Pereboom (2021b), Milam (2016), Sommers (2007), Shabo (2012, 2022), and Wolf (1981). Watson's (2014) reconstruction of what he takes to be Strawson's central argument ("the Normative Framework Argument") also hinges on something like this connection.

invites” and which Strawson rejects (2008c, p. 16f.). Interpretatively, for these reasons, Coates’s reading is implausible.⁴⁴

Furthermore, while Coates’s reading is offered as an alternative to the “relentlessly naturalistic” Standard Reading, Coates (2017, p. 818) does nevertheless attribute a form of naturalism to Strawson. The way Coates treats naturalism, it remains complementary to, indeed fundamental to, the transcendental argument: it fixates our antecedent commitment, *q*.

It is also a bit unclear what exactly “naturalism” means on Coates’s reading. Does it imply that our engagement in, or the value of, ordinary interpersonal relationships is *justified*? Different formulations suggest different renditions of this claim (cf., De Mesel, 2018). The same dialectic which suggests that the argument is a deductive transcendental argument—that it is not “really credible” that engagement in the practice of ordinary interpersonal relationships is unjustified, but “instead more plausible” that such engagement is justified; that it is “on secure ground”, as is “widely agreed” (Coates, 2017, p. 812)—also suggests that “naturalism” *justifies* belief, commitment, or engagement. Explicitly, according to Coates, Strawson’s naturalism means that “our friendships and love relationships [...] are *legitimate*” (2017, p. 818, emphasis added).⁴⁵ That we are so committed is a natural (psychological?) fact about us, in virtue of which the practice is justified.⁴⁶ If this is right, then Coates’s reading ultimately appears no less “relentlessly naturalistic” than the Standard Reading. The central motivation for his alternative reading is thus lost.

⁴⁴ If Strawson did employ deductive transcendental arguments in *Individuals* and only later changed his mind on their promise, favouring either elenctic transcendental arguments or no transcendental arguments but ‘the naturalistic response’ to scepticism instead, then it may be that this criticism of Coates’s reading as an interpretation of ‘Freedom and Resentment’ is flawed. But, as I argue in this chapter, there is no relevant methodological discontinuity to invoke.

⁴⁵ Sometimes, Coates’s seems to present Strawson’s “naturalism” as simply affirming (what is sometimes called) a *Moorean fact*, “one of those things that we know better than we know the premises of any philosophical argument to the contrary” (Lewis, 1996, p. 549). However, see Coates (2017, n. 17). If that is the sense in which Coates construes Strawson’s view, an interpretative problem is that Strawson contrasts his own view with Moore’s (1925, 1939), which he thinks, with Stroud (1979), entirely “misses the point of the sceptical challenge” (Strawson, 2008c, p. 5).

⁴⁶ That Strawson would be ‘a naturalist’ about something less contentious than the practice of responsibility would not make naturalism any less “relentless” or implausible, if naturalism is taken to mean that some natural necessity is itself to serve as a rational justification.

IV. Answering Glock's Question

That Strawson developed transcendental arguments against scepticism is a historical fact about 20th century analytical philosophy (see, Glock, 2003a, 2012, 2017; Peter M. S. Hacker, 2001; Stern, 2019, sec. 1). But the kind of transcendental arguments Strawson employed were not deductive (or otherwise directly anti-sceptical) transcendental arguments. If 'Freedom and Resentment' embodies or is based on a transcendental argument against the responsibility sceptic, this is—*pace* Coates's reading—an elenctic (or otherwise indirectly anti-sceptical) transcendental argument.

However, the philosophical program most often associated with 'Freedom and Resentment' is not that of neo-Kantian transcendental arguments, but that of naturalism. What exactly is the relation between the two anti-sceptical strategies in Strawson's methodology?

Return to our initial interpretative issue:

Glock's Question: What is the nature of the sceptic's mistake, according to Strawson?

Is the sceptic, and the responsibility sceptic in particular, "merely paralyzed in his thinking and acting? Or is there also an intellectual flaw in his position?" (Glock, 2022, p. 448). Is there an elenctic transcendental argument against the responsibility sceptic or is the responsibility sceptic's mistake not of this kind, but of the kind that a naturalist response to scepticism purports to demonstrate?

With respect to the set of apparently conflicting claims we've considered, and the different philosophical approaches producing these claims, one explanatory route might be, as is the suggestion of Robert Stern (2003), to ascribe a "naturalistic turn" to Strawson's philosophical development.⁴⁷ Stern argues that *Scepticism and Naturalism* marks a turn in the development of Strawson's thought, away from the Kantianism and the transcendental arguments of, in particular, *Individuals* and *The Bounds of Sense* and towards a Humean naturalism that propounds the inescapability of our natural disposition to belief in that which the sceptic purports to doubt. This suggestion partitions the seemingly conflicting characterisations of the sceptic's mistake to different *periods* in the development of Strawson's thought.

But we should not take this explanatory route—should not postulate a "naturalistic turn" or take this to answer Glock's Question—for the following reasons. First, even if there would have been a turn (which there wasn't),

⁴⁷ On this alternative, see Heyndels (2019, p. 412f.; 2020, p. 134f.) and Glock (2022, p. 446ff.).

Strawson's clarification that "the demand for justification is really senseless"—that it is not (in Glock's words) merely "of a psychological kind, as a naturalist gloss would have it"—comes *after Scepticism and Naturalism*; thus, on the supposed dialectic, this would have to count as a *return* (a suspiciously unannounced such).

Second, while it might not be dialectically felicitous at this point, it is generally assumed that 'Freedom and Resentment' expresses a naturalistic approach to the issue of responsibility and free will. This is obviously difficult to square with a "naturalistic turn" that is supposed to have materialised some 20 years after that paper and, importantly, after (the rather Kantian) *The Bounds of Sense* (which 'Freedom and Resentment' predates).⁴⁸

Third, in *Scepticism and Naturalism*, Strawson says that it is "the *naturalist* philosopher [that] will embrace the real project of investigating the connections between the major structural elements of our conceptual scheme" (2008c, p. 17, emphasis added; see also p. 18). The recommended project *for the naturalist* is Strawson's idea of descriptive metaphysics, which he presented already in *Individuals*—the project, that is, "to describe the actual structure of our thought about the world" (Strawson, 1959, p. 9, emphasis added). This is continuous with *connective analysis*, which Strawson consistently practiced. (We might want to say that what Strawson urges us to do as naturalists is connective analysis. However, as the focus on the "major structural elements of our conceptual scheme" suggests,⁴⁹ in this passage, what we may at most say is that Strawson urges the naturalist philosopher to embrace connective analysis-cum-descriptive metaphysics.⁵⁰) Whatever

⁴⁸ Callanan (2011, p. 276), Heyndels (2019, p. 412; 2020, p. 134), and Glock (2022, p. 446) also make this point. Furthermore, 'Social Morality and Individual Ideal' (2008d), Strawson's other significant paper in moral philosophy, also exhibits what we should recognise as naturalistic ideas—and, again, the two were written "about the same time", between 1960–1962 (Strawson, 1998a, p. 11). For some illuminating remarks on 'Social Morality and Individual Ideal', in connection to 'Freedom and Resentment', see De Mesel & Cuypers (2023).

⁴⁹ A focus that may stem merely from the dialectical context of the claim, i.e., regarding the role of transcendental arguments.

⁵⁰ On the relation between descriptive metaphysics and connective analysis, Strawson says that the former "does not differ in kind of intention" from the latter, "but only in scope and generality" (Strawson, 1959, p. 9). Hence, I glossed descriptive metaphysics, following Heyndels (2020, p. 23), as being *continuous* with connective analysis. Martinich (2024, p. 222) proposes that "connective analysis can be the goal of metaphysics", seemingly taking the relation to be rather the other way around. As the editors of a recent anthology on Strawson's work notes, the relation between the method of descriptive metaphysics and that of connective analysis "remains somewhat elusive" (Heyndels et al., 2024a, p. 7). For

naturalism (*à la* Strawson) comes to, as presented in *Scepticism and Naturalism*, it is not only *compatible* with Strawson's philosophical program (at least) from *Individuals* and forward, but is, furthermore, presented as being in some sense *behind* or *underlying*, as *motivating* or *grounding*, the project of connective analysis-cum-descriptive metaphysics, which he consistently propounded throughout his philosophical career.

Actually, explicit avowal on Strawson's part of naturalism traces back further than *Individuals*; this makes for a fourth reason against ascribing a "naturalistic turn" to Strawson. In *Introduction to Logical Theory* (1952, ch. 9)—before *Individuals*, 'Freedom and Resentment', and *The Bounds of Sense*—Strawson argued—much like he does about our practices of responsibility in 'Freedom and Resentment' (2008a, n. 7)—that inductive reasoning cannot be given a general justification.⁵¹ In response, Wesley Salmon (1957) objected that, by eschewing the reality of, need for, and search of a general justification of the practice of inductive reasoning, Strawson presents inductive reasoning as *merely conventional*; as it being, ultimately, a matter of *arbitrary choice* whether to reason inductively. Strawson's response is very instructive: he insists that the impossibility of providing a general justification of induction does not imply that it is a matter of *choice* whether to engage in inductive reasoning; there is, on Strawson's view, no choice here since induction is "forced upon us by Nature" (Strawson, 1958, p. 21). As others have noted (Glock, 2022, p. 446; Heyndels, 2019, p. 412; 2020, p. 134), this strongly tells against a "naturalistic turn": Strawson's reply to Salmon makes for a clear expression of a naturalistic way with the sceptic, by all appearances a response of the same kind as that which Strawson later elaborated on and made more explicit in *Scepticism and Naturalism*, but which predates the key works of his supposedly more Kantian phase.⁵² (If we are counting, this makes for five unannounced turns so far.)

For the reasons gathered here, we should not make sense of the apparent oscillation on Strawson's part between significantly disparate characterisations of the sceptic's mistake by postulating a "naturalistic turn".

Another way of making sense of the apparent oscillation between anti-scepticism in the form of naturalism and in the form of transcendental

relevant discussions, see Heyndels (2019; 2020, ch. 1), Glock (2012, 2017, 2022) and, of course, Strawson (esp., 1963; 1992a, ch. 2; 2011a, 2011b).

⁵¹ Strawson's claim is that "induction cannot be given a general justification, since 'to question induction in general leaves no canons in terms of which the justification can occur'" (1958, p. 20).

⁵² Cf., for example, Strawson's (1998d) reply to Putnam (1998).

arguments—i.e., of Glock’s Question—is Glock’s (2022, p. 449) own exegetical proposal: that Strawson *vacillates* between the different characterisations of the nature of the sceptic’s mistake and, correspondingly, between the anti-sceptical approach of naturalism and that of transcendental arguments.

But this also misrepresents Strawson’s view. His approach is not (exclusively) *either* a naturalism *or* to produce transcendental arguments. The two are not competing alternatives, but complementary anti-sceptical lines of argument.

Most succinctly put, this is Strawson’s answer to Glock’s Question:

scepticism is *at worst* senseless, *at best* idle. (Strawson, 1992a, p. 96, emphasis added).

While Strawson’s naturalism is more fundamental, it admittedly packs less anti-sceptical punch. It is worse for the sceptic if their challenge is “senseless”, but in any case, by the naturalist’s lights, their challenge is still “idle”. Transcendental arguments are not *instead of* a naturalistic response, but are potentially complementary to naturalism.

If connections as tight as those which transcendental arguments [...] claim to offer are really available, *so much the better*” (Strawson, 2008c, p. 8, emphasis added).⁵³

If anti-sceptical transcendental arguments can be successfully produced, then scepticism is for the worse; the sceptical challenge can then be demonstrably shown to be senseless.

One explanation for why the relation between these two anti-sceptical approaches has been misrepresented might be the seemingly paradoxical nature of seeing Strawson’s anti-scepticism as *both* universal *and* discriminating. On the two solutions to Glock’s Question considered above, the implicit assumption is that Strawson has a (more or less) universal anti-sceptical approach: *any* scepticism may be shown to be unfounded, *either* in the sense that it is “idle” *or* that it is “senseless”. This assumption itself is not incorrect. What is incorrect is the further assumption implicit in these readings: that a universal approach *excludes* a more discriminating way with sceptical

⁵³ Glock (2022, p. 449) argues that the transcendental strategy is preferable to the naturalistic strategy (also, Glock, 2012, pp. 413–417; 2017, pp. 224–226). Strawson would agree. The issue here concerns only whether the two strategies are excluding each other or exclusive of each other. The claim is that they are neither and that Strawson did not think they were.

challenges. While it is not always possible, when it is, Strawson dictates: produce a transcendental argument to demonstrate the sceptic's confusion.

Another explanation for why the relation between Strawson's naturalism and his employment of transcendental arguments, as his anti-scepticism more generally, has been misunderstood might be that Glock's Question puts us on slightly the wrong tracks. To ask, 'What is *the nature* of the sceptic's mistake, according to Strawson?' may unnecessarily pin the two approaches against each other, taking there to always be one specific and unified nature to the sceptic's mistake that always calls for the same response. We are better off asking: For each of the two responses, *under what circumstances* is that argumentative strategy appropriate as a response to some kind of scepticism? We may then see that Strawson is not *sometimes* a naturalist and *sometimes* a neo-Kantian employing transcendental arguments. He is *always* a naturalist.⁵⁴ Sometimes, however, rather than opting (simply) for "a naturalistic response" to scepticism, he produces a transcendental argument against the sceptic. Transcendental arguments are (dialectically) contentious (Gomes, 2017) and hinge on the strength and clarity by which the transcendental claim can be demonstrated.

V. Naturalism

Having covered the ambitions and limitations of the kind of transcendental arguments that Strawson employed, we are now left with the question: When a transcendental argument is not available, what is the anti-sceptical implication of what Strawson calls "the way of Naturalism" (2008c, p. 8)?⁵⁵

Our current interest lies with naturalism's possible anti-sceptical implications. The claim to be considered is whether naturalism may make for a universal anti-sceptical approach, a sweeping strategy that applies (more or

⁵⁴ Which does not exclude that he is also always a kind of neo-Kantian (albeit, obviously, not always deploying transcendental arguments).

⁵⁵ One way of elucidating Strawson's naturalisms is to prefix "naturalism" in different ways. For example, Glock (2022) variously characterises Strawson's naturalism as an "anthropological" naturalism and a "non-revisionary" naturalism. Paul Russell's (2017d) classic discussion of Strawson naturalism adjudicates between understanding him as a "type-naturalist" or a "token-naturalist". Strawson (2008c, ch. 1) himself, when most explicit, stresses that it is a "catholic", "liberal" and "non-reductive" naturalism, as opposed to a "strict" or "reductive" naturalism that he espouses. These labels are variously helpful, and we will discuss some of them in more detail further on.

less) in the same way to every kind of philosophical scepticism; a method that is more fundamental than the method of transcendental arguments, but at least compatible with and possibly congenial to such arguments.

First of all, above we glossed “the naturalistic response” (following Glock’s presentation) as taking the sceptic’s mistake to be “of a merely psychological kind” such that the sceptic is “merely paralyzed in his thinking and acting” (Glock, 2022, p. 448). It is very commonly assumed that, for example (and in particular), the reactive attitudes are inescapable in the sense that they are *psychologically necessary*, or too deeply *psychologically entrenched* in our mental or emotional constitution for them to be assailable by theoretical arguments. Contrary to this common conception, this is not the sense in which, for that which is inescapable, sceptical doubt is “idle” or “in vain”, according to Strawson. The Inescapability Claim is indeed a central aspect of Strawson’s naturalistic approach. The sense in which some aspects of our conceptual scheme are inescapable is an important interpretative question—not least because the Standard Reading is implausible, both interpretatively and independently. At this point, however, while what is said below is highly relevant for this issue, I will only assert that it is in any case *not* a matter of *psychological* necessity or entrenchment and defer further engagement with this central question until chapter 5.⁵⁶

The aspect of Strawson’s naturalism to focus on now is the claim that rational justification and doubt are *internal to a framework*, and that without this framework and with respect to the framework itself, neither justification nor doubt makes sense.⁵⁷

⁵⁶ For a relevant discussion of ways in which beliefs can be said to be *basic*, from which I adopt the notion of psychological entrenchment, see Glock (2016, p. 275ff.)

⁵⁷ Heyndels (2020, *passim*) calls this aspect of Strawson’s naturalism “[Natural Framework]”, distinguishing it from two further aspects: “[Inescapable]” and “[Not A Belief]”. For discussion, see Glock (2022, p. 447f.). Joe Campbell (2017) have persuasively argued that several prominent readings of Strawson on free will and responsibility mistakenly ascribe to him such a foundationalist justificatory aim. There is here a crucial distinction—illuminated by Strawson’s recurrent reference to Wittgenstein (especially, *On Certainty* (1969)) with respect to this idea (e.g., Strawson, 1998d; 2008c, p. *passim*)—between, on the one hand, what may be justified or unjustified and, on the other hand, what is *beyond being justified or unjustified*. Moves within the framework—for example, assertions or doubts that are *of* the framework in the sense that they are the product of that framework, presupposing the framework if they themselves are to be meaningfully expressed—may be correct or incorrect, justified or unjustified. The framework itself, however, according to Strawson, is *beyond justification*. Recently, Rummens and De Mesel (2023) have argued that moral responsibility is a basic certainty, beyond justification, and that this seems also to be Strawson’s view. We will come back to this in chapter 5.

In ‘Freedom and Resentment’, this idea is most clearly expressed in the following passage:

Inside the general structure or web of human attitudes and feelings of which I have been speaking, there is endless room for modification, redirection, criticism, and justification. But questions of justifications are internal to the structure or relate to modifications internal to it. The existence of the general framework of attitudes is itself something we are given with the fact of human society. As a whole, it neither calls for, nor permits, an external ‘rational’ justification. (Strawson, 2008a, p. 25)

On Strawson’s account, *this* is what “[p]essimists and optimists alike show themselves, in different ways, unable to accept”; why they “over-intellectualize the facts” (2008a, p. 25). The mistake that is “in common to their misunderstandings”—and, importantly, as he notes, common also to the mistake of “the genuine moral sceptic” (2008a, p. 25)—is the (implicit) endorsement of a *foundationalist* picture of our conceptual practices.

Strawson’s naturalism—pivoting on the idea of a framework delineating rational competence—is, as an anti-sceptical approach, an *anti-foundationalism*; and Strawson’s naturalism is, centrally, anti-sceptical. It is in these negative terms that we get the clearest expression by Strawson of what his naturalism is. Discussing how we should view anti-sceptical transcendental arguments, he says: we should not construct a counter-argument to scepticism that aims at “supplying the reasoned rebuttal which the sceptic perversely invites”; we should not, because “naturalism *is precisely the rejection of that invitation*” (Strawson, 2008c, p. 16f., emphasis added). Strawson’s naturalism is, centrally,

the [...] rejection both of scepticism and of scepticism-rebutting arguments as equally idle—as both involving a misunderstanding of the role in our lives, the place in our intellectual economy, of those propositions or crypto-propositions which the sceptic seeks to place in doubt and his opponent in argument seeks to establish. (Strawson, 2008c, p. 16)

From this characterisation we get both that Strawson’s naturalism is at its core *anti-sceptical*—a rejection of that perverse invitation—and that it is anti-sceptical in virtue of being an *anti-foundationalism*—an abandonment of “the unreal project of wholesale validation” of our conceptual scheme (2008c, p. 17).⁵⁸

⁵⁸ It is mistaken, therefore, to claim as Nicholas Sars (2022a, p. 79) has recently done, that Strawson “thinks there is a perspective external to our given framework of attitudes from

We also see, with respect to ‘Freedom and Resentment’, that while “the genuine moral sceptic” may seem like a peripheral character (mentioned only twice), he is the central target of the approach. The genuine moral sceptic—correctly, by Strawson’s lights—“sees that the optimist’s account is inadequate and the pessimist’s libertarian alternative inane” (Strawson, 2008a, p. 25). If we stay committed to “the unreal project of wholesale validation”—i.e., the foundationalist picture of our conceptual scheme—scepticism is, in a sense, correct: it reflects due recognition of the failure of that project. But scepticism isn’t correct. Because scepticism, just as the other parties in the debate, fail to recognise that the framework itself “neither calls for, nor permits, an external ‘rational’ justification” (Strawson, 2008a, p. 25). For just this reason, any acceptable optimism about responsibility must be free from the commitment to the foundationalist project—this is the *radical* modification Strawson seeks to effect (2008a, p. 27). Only then can we properly recognise that “the facts as we know them supply an adequate basis for the concepts and practices which the pessimist feels to be imperilled by the possibility of determinism’s truth” (Strawson, 2008a, p. 2). To “sufficiently, that is, *radically*, modify the view of the optimist” (Strawson, 2008a, p. 27) is to purge it from the foundationalist picture.

Does naturalism, so far understood, supply the formula for a *universal* anti-sceptical approach? Foundationalism, as scepticism, does take somewhat different forms with respect to different issues—compare epistemic foundationalism (and scepticism) with moral foundationalism (and scepticism).⁵⁹ If we are to make sense of Strawson’s naturalism as a universal

which the entire framework itself might be evaluated”. Sars’s reason for thinking this is that “Strawson [...] acknowledges the possibility (at least in principle) of ‘a question about the rational justification of ordinary inter-personal attitudes in general’ [(Strawson, 2008a, p. 14)]”. But it is exactly *in principle* that he denies it. In the sentence immediately following the one Sars cites, Strawson says “such a question could seem real only to one who had utterly failed to grasp the purport of the preceding answer, the fact of our natural human commitment to ordinary inter-personal attitudes” (Strawson, 2008a, p. 14).

⁵⁹ ‘Foundationalism’ perhaps most commonly refers to *epistemic* foundationalism, which holds that the structure of epistemic justification requires, for one thing, some beliefs that are foundational in the sense that they are justified and their justification is unconditional on other beliefs, and, furthermore, that any non-foundational beliefs depend, ultimately, on foundational beliefs if they are to be justified (Hasan & Fumerton, 2022). The idea is that *conceptual foundationalism*—i.e., foundationalism about our conceptual scheme—retains basically the same structure: there must be foundational reasons for having these concepts and our reasons for concept-use derive their legitimacy, ultimately, from these foundational reasons. This formulation, however, captures only, what we may call, *rationalist* conceptual foundationalism, and not, what we may by way of contrast call, *empiricist* conceptual foundationalism. (If to articulate this version, Quine’s ‘naturalizing’ project may be a source of inspiration.) Conceptual anti-foundationalism amounts, if not to conceptual relativism (as

anti-sceptical approach, as Strawson himself views it, then we do best in understanding his anti-foundationalism as applying most generally. Let's take the target of Strawson's anti-foundationalism to be *conceptual foundationalism*: the view that a rational foundation is necessary for our conceptual scheme. The rational foundation may be sought in, for example and saliently, human nature or universal reason (or, as are other traditional options in ethics, divine commands or natural law).⁶⁰

The question the sceptic poses is: What is the rational justification for employing this conceptual scheme, the one that we do employ (including, in particular, the feature of it which the sceptic challenges), rather than some other conceptual scheme (without this feature)? Scepticism is not only the view that we lack a rational justification for employing our present conceptual scheme. If this was all the sceptic claimed, Strawson could be said to agree with the sceptic: there is no such justification. Rather, the sceptic holds in addition that there *need* to be a rational justification of our conceptual scheme. This is where the anti-foundationalist, Strawson, disagrees: the absence of a rational foundation is no flaw—"there can only be a *lack* where there is a *need*" (Strawson, 2008c, p. 33, original emphasis). The naturalist recognises the framework as itself delineating our rational competence—such that any question of justification and any doubt is internal to and presupposes the framework, while the framework itself is not open to either justification or doubt—and finds *no need* for a rational foundation of the framework itself.

This central feature of Strawson's naturalism explains, in a unifying way, the passages considered at the beginning of this chapter, which seemed to intensify the disparity of Strawson's different characterisations of the sceptic's mistake. We need to observe that while the predicated mistake in each passage is different, *the subject* of predication is also different between the passages. One passage takes as its subject "sceptical arguments and rational counter-arguments". The other passage takes as its subject "the demand for justification".

In presenting naturalism as an alternative to transcendental arguments, the point is to see that the sceptical arguments as well as any rational counter-

it does not in the case of Strawson), at least to the combination of two ideas: one the one hand, that language is autonomous, or grammar arbitrary—i.e., that our conceptual scheme cannot itself be correct or incorrect, at least not with reference to some putative essence of reality—and, on the other, that there is no reason for why we have the conceptual scheme we have. See, in particular, Strawson (2008c, p. 14), for him siding with Wittgenstein against Carnap on the latter point. For a concise discussion of further features of the foundationalist project, see Michael Williams (2005).

⁶⁰ I adapt this gloss of foundationalism from Queloz & Cueni (2021, p. 2).

arguments (i.e., direct responses to scepticism) are *idle*. While *the point* here is not to show that they are senseless, this does not mean that a sceptical challenge cannot be (even if they may not always be shown to be) senseless. What is always senseless, however, is the demand for justification. Naturalism does not say that the lack of a foundational justification is no flaw of our conceptual scheme because we simply cannot help but to have this scheme. The lack of a foundational justification is no flaw because the demand for justification is *senseless*: to take us to need such justification shows a misunderstanding—a mistaken picture; a foundationalist understanding—of our conceptual scheme. While it may seem like Strawson contrasts two different characterisations of the sceptic’s mistake, and that he is oscillating between them, the different characterisations pertain to different mistakes, and both are rendered evident as mistakes from the perspective of the kind of naturalism that Strawson champions.

That, however, is not to say that there *are* not two different mistakes. There are; we will consider this matter in chapter 5. The point here is that the different characterisations in the different passages are not *conflicting* characterisations of the sceptic’s mistake.

VI. Conclusion

Strawson does not meet any and all sceptical challenges with a transcendental argument, simply because he does not think that all such challenges can be so met. The transcendental argument is indeed an effective anti-sceptical method, but the (dialectically felicitous) demonstration of a transcendental claim demands much, not just by way of insight and imagination on the part of the argument’s proponent, but also, and more fundamentally, of the actual tightness of the conceptual connections concerned. Therefore, Strawson sometimes takes recourse to a different response: “the way of Naturalism”.

Regarding the anti-sceptical import of naturalism, the claim is that it deflates scepticism by showing how it relies on a foundationalist picture of our conceptual practices. This is intended as a *universal* anti-sceptical method. But it does not in any way exclude the possibility that, *when* “connections as tight as those which transcendental arguments [...] claim to offer are really available”, *then* we may show the particular sceptical challenge (itself) to really be *senseless*.

Beside the various interpretative issues that have occupied us in this chapter, circling Glock’s Question, there is also the independent, substantive issue of

whether Strawson's way with scepticism (about any particular issue) is convincing. Our question is, in particular: is his way with scepticism *about responsibility* convincing? There is also the further, related issue—just as substantive but less independent from interpretation—of if (and if so, how) we can make good on Strawson's Promise.

I have myself a vague worry about the attitude Strawson's approach may encourage; perhaps we should qualify Strawson's position slightly, rendering the claim less grand but more plausible, by adding that *if (or in so far as)* scepticism is encouraged by or premised on a foundationalist picture of our conceptual practices, then (or to that extent) it is, by the light of naturalism, mistaken. Strawson may be right that philosophical scepticism has traditionally been beholden to this picture. But the qualification presents it as an open question if (or to what extent) this is actually the case for any sceptical challenge; it leaves it open whether a serious challenge of our conceptual scheme (or a particular framework feature of it) could not be mounted in the anti-foundationalist's own terms.

If we only focus on the prospect of making good on Strawson's Promise by a rational reconstruction of his approach, what are we to say at this point? We early on noted that the naturalist response to scepticism is not well-regarded, not even among those who profess to be otherwise sympathetic to Strawson's approach to responsibility. This owes greatly to a common misunderstanding regarding what that response even is. Nevertheless, before we turn to clear some of the most unfortunate misunderstandings (in chapter 5), it will be instructive to consider some alternative ways in which readers of Strawson has sought to vindicate what has seemed, and to some extent still does seem, so promising about Strawson's approach to responsibility.

In the next chapter, we will consider Pamela Hieronymi's (2020) novel interpretation of Strawson's approach in 'Freedom and Resentment'. She takes the central argument of that paper to rest on a transcendental argument (Hieronymi, 2020, p. 28). For this reason, it should already be noted, first, that Coates's reading was found to be interpretatively implausible, centrally, because it ascribed *the wrong kind of* transcendental argument—indeed, more generally, the wrong kind anti-sceptical approach—to Strawson, not because it ascribed *a* transcendental argument to him; second, that if we thought (as some seem to have thought) that Strawson's transcendental strategy somehow conflicts or competes with his naturalist response, then, by revealing that thought as unfounded, the way has in fact been paved for a reading that finds a transcendental argument in or behind 'Freedom and Resentment'.

3. The Incoherence Puzzle

I. Introduction

Pamela Hieronymi's recent book *Freedom, Resentment, and the Metaphysics of Morals* (2020) starts with puzzlement. Her novel reading of Strawson's approach to the problems of free will and responsibility in 'Freedom and Resentment' gains its initial impetus from what she sees as an explanatory gap in the interpretative landscape. Her claim is that, in light of the existent interpretations, 'Freedom and Resentment' is (and ought to be) *puzzling*: none of the existent interpretations "would lead you to expect what you will find, looking at the central text" (2020, p. 2). What we supposedly find if we look at what Hieronymi considers "the central argument" is that Strawson charges his opponent with *incoherence* or *self-contradiction*. The question is: Why does he do that? What is the argument to that conclusion? This is the source of puzzlement: for any of the existent interpretations, so Hieronymi finds evident, this charge of incoherence on Strawson's part remains an unaccounted for and perplexing aspect of the paper; all other readings simply fail to make sense of it. The prime motivation for Hieronymi's novel reading of Strawson, that is, is explicitly to construct an account capable of explaining this neglected interpretative issue.

This is the interpretative puzzle:

The Incoherence Puzzle: Why does Strawson charge his opponent with incoherence or self-contradiction?

We will here assume, with Hieronymi, that there is an interpretative puzzle to be addressed. This assumption, as we will come to see, is not wholly unproblematic. For, while trying to solve the Incoherence Puzzle, Hieronymi's reading, ironically, provokes two further interpretative puzzles: *the Second Incoherence Puzzle* and *the Self-Contradiction Puzzle*. The Second Incoherence Puzzle concerns the fact that Strawson accuses his opponent with incoherence *only once*. We will come to see that this is, especially for Hieronymi's reading, surprising. A very natural response to this on

Hieronymi's part is, as she herself argues, that Strawson as a matter of fact, although less explicitly, accuses his opponent of incoherence *twice*—not once. However, this retort—even if it were correct, which it is not—nevertheless fails to address what is puzzling about the Second Incoherence Puzzle. The further puzzle provoked, the Self-Contradiction Puzzle, is that, on Hieronymi's reading of Strawson and on her solution to the (initial) Incoherence Puzzle, Strawson appears to be flatly contradicting himself. In a previously unpublished private letter, Strawson denies a central presupposition of the view which Hieronymi ascribes to him. In spelling out the argument that Strawson's opponent is committed to a self-contradiction, Hieronymi has inadvertently committed Strawson himself to a self-contradiction.

The Self-Contradiction Puzzle does not spell trouble only for Hieronymi's novel reading. It also unsettles the Conventional Way with "the central argument" (and, by extension, with the Incoherence Puzzle), since the Conventional Way ascribes to Strawson the very same problematic presupposition which commits him to a self-contradiction. Hieronymi's injunction to study the text closer proves itself, somewhat incidentally, to be not without consequence for our ordinary understanding of the text.

Thus, neither the established way with "the central argument" nor the novel way with that argument which Hieronymi presents satisfactorily deals with the Incoherence Puzzle. Beside the substantive interpretative issues represented by the Second Incoherence Puzzle and the Self-Contradiction Puzzle, its failure to address the Incoherence Puzzle straightforwardly undermines the prime motivation for Hieronymi's novel reading. But since neither the Conventional Way with the "the central argument" is satisfactory, we cannot leave it on this negative note; some puzzlement about Strawson's approach should indeed remain with us. Probing the text, we gather before us an agglomeration of interpretative puzzles—the Incoherence Puzzle, the Second Incoherence Puzzle, and the Self-Contradiction Puzzle—but, I wish to contend, we are not for the worse therefore. In closing, I offer some reflections on how we may deal with the interpretative situation which this surge of further and further puzzles may now, somewhat paradoxically, have helped us to see more clearly.

II. The Statistical Reading and the Conventional Way

Hieronymi presents a novel and avowedly unconventional interpretation of Strawson's approach in 'Freedom and Resentment'. As noted, the prime

motivation for her reading is that it can account for a heretofore puzzling feature of ‘Freedom and Resentment’, while no other existent interpretation can. More precisely, on Hieronymi’s presentation there are broadly speaking two different readings of Strawson’s approach: ‘the simple Humean interpretation’—that, “given our psychological limitations, we are struck treating one another *as if* we are morally responsible”⁶¹—and ‘the broadly Wittgensteinian interpretation’—that the practice itself cannot be given a justification, as issues of justification are internal to the practice. On neither of these would we anticipate a charge of incoherence or self-contradiction against the sceptic on Strawson’s part (Hieronymi, 2020, p. 2, also p. 59). Nevertheless, that is what we find. Thus,

The Incoherence Puzzle: Why does Strawson charge his opponent with incoherence or self-contradiction?

To make interpretative sense of this, Hieronymi provides a novel reading of Strawson’s approach; let’s call it, *the Statistical Reading*. The Statistical Reading earns its name from the way it proposes that we understand “the central argument”. First, consider the argument as Strawson presents it:

The Central Argument: “I remarked that the participant attitude, and the personal reactive attitudes in general, tend to give place, and it is judged by the civilized should give place, to objective attitudes, just in so far as the agent is seen as excluded from ordinary adult human relationships by deep-rooted psychological abnormality – or simply by being a child. *But it cannot be a consequence of any thesis which is not itself self-contradictory that abnormality is the universal condition*” (FR: 12, emphasis added)

Hieronymi thinks that there is a straightforward way of understanding this argument on which it at least makes sense, is “a worthy contender” (2020, p. 3), or is, even, “a powerful argument” (2020, p. 16, also, pp. 2, 71). The straightforward way of understanding the argument is by understanding “abnormality” (and correspondingly, “normality”) *in a statistical sense*. If we so understand it, the idea is, then “[t]he conclusion follows immediately” (Hieronymi, 2020, p. 17): *everyone cannot be statistically abnormal; most people must be statistically ordinary*.⁶² This answers Strawson’s opponent (the

⁶¹ That is, a reading that treats the Inescapability Claim as the central argument for Strawson’s approach and understands this claim as on, what we dubbed, the Standard Reading.

⁶² Hieronymi’s claim is actually stronger than that this is *a* way, or a particularly straightforward, intuitive or otherwise good way, of making sense of the argument; rather, she claims: “I do not see how to make sense of the passage without it [i.e., without the assumption that

responsibility sceptic) because what the opponent is here taken to suggest is that there might be an exempting condition—say, that the agent could not have acted otherwise—which may hold true for each and every individual human agent, such that everyone is in fact exempted and no-one is in fact responsible. That is, Strawson is here taken to prevent what has come to be known as *the generalisation strategy*.⁶³

It is worth considering Hieronymi's own presentation of Strawson's argument:

Notice that Strawson's seemingly facile argument [i.e., "the central argument"] starts with a claim—a very strong, questionable claim—about the reason why we exempt, why we do and should suspend the reactive attitudes: we do and should "just insofar as [that is, if, only if, and to the extent that [Hieronymi's clarification]] the agent is seen as excluded from ordinary adult human relationships." [(Strawson, 2008a, p. 12)] Strawson claims that we exempt people from these attitudes just in case we believe they are (as he later puts it) "incapacitated" for ordinary adult interpersonal relationships. And, from this single premise, he immediately concludes that his opponent is committed to a contradiction.

The conclusion follows immediately, if we interpret Strawson's "ordinary" as "statistically ordinary." On this interpretation, Strawson's first premise claims that we do and should suspend the reactive attitudes just insofar as we believe the agent is excluded from statistically ordinary interpersonal relationships. Strawson's opponent thinks that a general thesis—something true of everyone—will give us reason to suspend the reactive attitudes. So, by

'abnormal' means 'statistically abnormal']. Once we make this interpretative move, the rest of the paper follows" (Hieronymi, 2020, p. 18, n. 11). According to Hieronymi, the claim's being baffling may be the reason why "this interpretation of Strawson has not been previously considered" (Hieronymi, 2020, p. 23). If this means that it has not been *seriously* considered, that might be right. But it has been to some extent considered. Russell, making the critical point that "it is not abnormality, as such, that excuses but, rather, incapacity" (2017d, p. 42) may be understood as cutting against precisely something like the statistical interpretation of the premise. Also, McKenna and Pereboom explicitly consider that interpretation, "On its face, it appears that [Strawson] is making the point that abnormality is statistically at odds with applying to all cases", but they conclude that "Strawson should not be read in this way" even if they "grant that the text invites this reading" (McKenna & Pereboom, 2016, p. 145, n. 10).

⁶³ The generalization strategy starts from a case or several cases in which we take a person to not be responsible, purports to identify a common reason why it is appropriate to not take a person to be responsible in these cases, and then seeks to generalize this attitude to all cases by showing that this common reason always applies if determinism is true. For discussion, see, e.g., Wallace (1994, ch. 5-6).

Strawson's first premise, the opponent must think this general thesis will give us reason to believe that everyone is excluded from what is statistically ordinary—that everyone is abnormal. But that is a contradiction. And, any thesis which implies a contradiction must itself be contradictory, not a “coherent thesis” [(Strawson, 2008a, p. 11)].⁶⁴

This is the Statistical Reading of “the central argument”. There are three important assumptions on this reading of the argument. First, that the thesis of determinism is a completely *general* thesis—meaning, it is true of everyone, everywhere, always.

Second, that the reasons for which we do (and should) exempt agents from responsibility are those that show that agents are incapable of engaging in *ordinary* interpersonal relationships. On the Statistical Reading, what are ‘ordinary interpersonal relationships’ is determined by what is *statistically* normal, such that the capacities required for being a responsible agent, the capacities which regulate our excusing and exempting practices, will be those that are statistically normal, or, alternatively, those of the statistically normal person.⁶⁵ The reasons for which we exempt, then, will be such that only those that are *statistically abnormal* are exempted. If everyone is determined (in whatever sense that determinism could imply) their being determined cannot, on this account, count as a reason for exempting them from responsibility.

Strawson's ‘core idea’, according to Hieronymi (2020, p. 18, also p. 33), is that “nothing true of everyone could give us reason to suspend the reactive attitudes”. Those that we do exempt are “outliers” and “it could not be the case that we are all outliers” (Hieronymi, 2020, p. 23). Since the sceptic claims that everybody is exempted, but we exempt only those that are statistically abnormal, the argument of the sceptic is self-contradictory, proposing an incoherent thesis. One the assumption that this is Strawson's ‘core idea’, then

⁶⁴ We should be careful here not to equate responsibility scepticism (or hard determinism) with the thesis of determinism: these are two quite different theses. It is about determinism that Strawson says “if there is a coherent thesis of determinism, then there must be a sense of ‘determined’ such that, if that thesis is true, then all behaviour whatever is determined in that sense.” (Strawson, 2008a, p. 11). Plausibly, the self-contradictory thesis that would have it that abnormality would be the universal condition is *scepticism* (or hard determinism). It is the *interpretation* of what determinism would mean, according to which it would have such a consequence, that is self-contradictory—scepticism (or hard determinism), that is.

⁶⁵ Hieronymi is not clear on this point. Since the statistically normal person may be only a model of a person which has all and only the statistically most prevalent properties, thus representing the set of people at large without (necessarily) being identical with any individual member of that set, it is plausible to assume that Hieronymi is working with a different conception of statistical normalcy.

we have a way to make sense of “the central argument”; the accusation of incoherence is no longer puzzling.

The third assumption of Hieronymi’s reading, evident as it is, must still be explicitly discussed; and in connection, an alternative to the reading. Hieronymi takes “the central argument” to be *central* to Strawson’s approach in ‘Freedom and Resentment’. Naturally enough, on this assumption, the need to solve the Incoherence Puzzle becomes pressing. In a way, to say that the Incoherence Puzzle is pressing if we assume that the charge of incoherence is the *central* argument of ‘Freedom and Resentment’ is to undersell the importance of this assumption. For there is a sense in which the assumption itself creates the puzzle. Hieronymi is right to say that “the central argument”, just as it is stated, has not been given much attention and is “typically overlooked” (2020, p. 17). But perhaps that is because Strawson himself admits that, at least in a sense, the argument is “too facile”? Perhaps, not unrelatedly, the argument has not been given much attention because the general opinion is that it has been given *enough* attention? Indeed, “the central argument” was forcefully objected to in 1992 by Paul Russell, and it is right that, at least since then, it has not been treated with any particular interest and has at least not been considered convincing, and certainly not as “a worthy contender” (Hieronymi, 2020, p. 3).

An alternative to Hieronymi’s interpretative approach to ‘Freedom and Resentment’ is, what we will call, *the Conventional Way* with the incoherence puzzle, which treats “the central argument” as indeed an “altogether too facile” argument. On this understanding, there is nothing puzzling about it: it is just a bad argument. Russell’s own case to this effect charges Strawson with an equivocation between “abnormality” and “incapacity”; whereas the latter is what is relevant for whether to exempt, “the central argument” only follows—i.e., the opponent is only committed to an incoherent thesis—if we (fallaciously) take “abnormality” to be the relevant concept. On the Conventional Way, “the central argument” is a mistaken, misconceived and implausible argument; but, luckily for Strawson, the argument is not central to his approach, at least in the sense that there is much else to his approach that makes for a novel, forceful, appealing, or at least suggestive way with the issue of free will and responsibility.

Tentatively, we will not follow the Conventional Way but rather join in Hieronymi’s puzzlement and indeed follow her interpretative injunction to follow the text closely. When we do that, however, what we find are further and further interpretative puzzles.

III. The Second Incoherence Puzzle

The charge of incoherence is to be expected and is straightforwardly accounted for on the Statistical Reading.

But (to borrow a phrase from Strawson) there is something else which, because this is true, is equally certainly *not* true. And that is that the central argument of ‘Freedom and Resentment’ really is that scepticism is self-contradictory. *If* Strawson’s argument were that the sceptical suggestion is *incoherent*—as the Statistical Reading has it—then it should indeed be surprising that Strawson declares the sceptical suggestion incoherent *only once*. Enter: our second interpretative puzzle:

The Second Incoherence Puzzle: Why does Strawson charge his opponent with incoherence or self-contradiction *only once*?

Why is the fact that the charge of incoherence occurs only once puzzling? Because a central feature of the dialectical structure of Strawson’s argument is that section IV and section V of ‘Freedom and Resentment’ are largely analogous. Strawson seeks to make essentially the same argument in both sections: first in the *less* “disputant-crowded field” of the *personal* reactive attitude (section IV) and then in the *more* “disputant-crowded field” of the *vicarious analogues* of these attitudes, viz. the *moral* reactive attitudes (section V). In section V, for example, we find the following programmatic statement: “What concerns us [...] is to inquire, *as previously* in connection with the personal reactive attitudes, what relevance any general thesis of determinism might have to their vicarious analogues” and, as he makes clear, “*The answers once more are parallel*” (Strawson, 2008a, p. 19, emphasis added). Indeed, he flags this feature of the argument throughout section V:

“*as before*, that when the suspension of such an attitude or such attitudes occurs in a particular case, it is *never* the consequence of the belief that the piece of behaviour in question was determined [(in the relevant sense)]”; and,

that, just as “we cannot take seriously the thought that theoretical conviction of such a general thesis would lead to the total decay of the personal reactive attitudes”, we cannot take this thought seriously with respect to the moral reactive attitudes. (Strawson, 2008a, p. 19, first emphasis added)

Adding, furthermore,

that, if one presses the matter further, asking whether we *should* not nevertheless abandon the practice of responsibility, this person “has wholly failed to grasp the import of the preceding answer, the nature of the human commitment that is here involved”; and,

“add, *as before*, that if there were, say, for a moment open to us the possibility of such a godlike choice, the rationality of making or refusing it would be determined by quite other considerations than the truth or falsity of the general theoretical conviction in question” (Strawson, 2008a, p. 20, last emphasis added)

Despite the declared analogical structure of Strawson’s overarching argument, the manifest repetition, and the repeated remarks signalling this repetition—in section V, a charge of incoherence is nowhere to be found. That section IV and section V are to be largely analogous is a fact about ‘Freedom and Resentment’. The charge of incoherence, however, makes for a disanalogy.

After all, then, it is not the case, *pace* Hieronymi’s ambition, that the Statistical Reading “would lead you to expect what you will find, looking at the central text”—viz., a *disanalogy* between section IV and V concerning the supposedly *central* argument. The Second Incoherence Puzzle is puzzling on the assumption that Strawson’s *central* argument, or in any case that *a* central argument, against the responsibility sceptic is that the sceptical challenge is incoherent. On that assumption, what we should expect is a second charge of incoherence. But that is not what we find. Hieronymi does not address this interpretative issue. The Second Incoherence Puzzle thus counts against the Statistical Reading. If we do not take seriously the (first and only) charge of incoherence, we of course avoid the Second Incoherence Puzzle. But without that assumption, the key motivation for Hieronymi’s novel interpretation disappears.

It might seem like we face a dilemma: if we take the charge of incoherence seriously and adopt the Statistical Reading, then we encounter the Second Incoherence Puzzle; if we do not adopt this reading, then the (first) Incoherence Puzzle remains unaccounted for. But that would be to overlook a salient alternative: the Conventional Way, which holds that “the central argument” is, as Strawson’s himself admits, “too facile” (2008a, p. 12). Notably, for the Conventional Way to fend off the Second Incoherence Puzzle, the sense in which the argument is “too facile” have to be such that “the central argument” can’t really be the *central*, or even *a* central, argument of ‘Freedom and Resentment’. Admittedly, if the (first) Incoherence Puzzle should not haunt us, the Conventional Way ought to supplement this non-centrality claim with an explanation of why Strawson nevertheless presents us with the argument in the

first place; why he, *at least initially*, charges the opponent with incoherence. On Russell's original discussion, the diagnosis of this is an *equivocation*. But this is somewhat surprising if it is, as Russell intends for it to be, a charge *against* Strawson, for it is Strawson who deems—indeed, is to deem, if the Second Incoherence Puzzle is to be circumvented—the argument “too facile”. Could *his* reason be that he equivocates? We will return to this later on.

IV. A Second Accusation of Incoherence? An Irrelevance

It may be thought that the Second Incoherence Puzzle is wholly unfounded, premised on a confusion regarding what the (initial) Incoherence Puzzle in fact is; indeed, premised on a lack of attention to what Hieronymi in fact presents us with. The *actual* motivation for surpassing the existent readings with the Statistical Reading is, as Hieronymi explicitly says:

[that] neither [of the existent] interpretation[s] would lead you to expect what you will find, looking at the central text: Strawson *twice* accuses his opponent of being caught in some kind of contradiction” (2020, p. 2, emphasis added; also pp. 43, 59).

The Incoherence Puzzle, it might thus be said, is not merely that Strawson charges his opponent with incoherence or self-contradiction but, more precisely, that he does so *twice*. What is the redeeming force of this observation? Does it show that there really is no Second Incoherence Puzzle?

It does not dissolve the Second Incoherence Puzzle. There are two reasons for this, each by itself sufficient to undermine the defence just mounted. First of all, even if Strawson had accused the sceptic of incoherence twice, as Hieronymi claims, the second time would still be—by Hieronymi's (2020, p. 43) own account—just as the first, in *section IV* of ‘Freedom and Resentment’. Thus, we would still lack an iteration of, and an analogue to, “the central argument” (where Strawson indeed does accuse the sceptic of incoherence), or, more generally, any charge of incoherence or self-contradiction, to be located somewhere in section V, which is what we should expect if the argument is indeed central (especially, if it is *the* central argument). Even if we grant to Hieronymi this second accusation of incoherence, the problem of explaining the disanalogy remains: the Second Incoherence Puzzle has not been resolved.

Second, Strawson does not in fact, *pace* Hieronymi's account, accuse the sceptic of being involved in incoherence or self-contradiction *twice*. What we find looking at the text that Hieronymi directs us to is not a charge of incoherence, but only a survey of the reasons for which we exempt some agents. There is no accusation of incoherence or self-contradiction to be found.

Of course, the charge of incoherence might be *implicit*; plausibly, that is just what Hieronymi would say in response. To address this possibility, we will first see that there is reason to think that the argument that is supposed to implicitly involve a charge of incoherence or self-contradiction is in fact, at least according to Strawson himself, an argument for the irrelevance of determinism (not, that is, for the incoherence of scepticism). In the next section, we will consider Hieronymi's own reconstruction of the relevant piece of text. This provides occasion for us to consider some central commitments of the Statistical Reading. As it will turn out, in light of previously unpublished material, the Statistical Reading of the relevant argument commits Strawson himself to self-contradiction; this is at least as puzzling as the other interpretative puzzles here considered.

First of all, what does Strawson actually say at the point where Hieronymi purports that there is a *second* accusation of incoherence?⁶⁶ He says that,

it is certainly true that in the case of the abnormal, [...] our adoption of the objective attitudes is a consequence of our viewing the agent as *incapacitated* in some or all respects for ordinary inter-personal relationships. He is thus incapacitated, perhaps, by the fact that his picture of reality is pure fantasy, that he does not, in a sense, live in the real world at all; or by the fact that his behaviour is, in part, an unrealistic acting out of unconscious purposes; or by the fact that he is an idiot, or a moral idiot. But there is something else which, *because* this is true, is equally certainly *not* true. And that is that there is a sense of 'determined' such that (1) if determinism is true, all behaviour is determined in this sense, and (2) determinism might be true, i.e. it is not inconsistent with the facts as we know them to suppose that all behaviour might be determined in this sense, and (3) our adoption of the objective attitude towards the abnormal is the result of a prior embracing of the belief that the behaviour, or the relevant stretch of behaviour, of the human being in question *is* determined in this sense. (Strawson, 2008a, p. 14, original emphasis)

⁶⁶ There is a second *explicit* invocation of "incoherence" (or self-contradiction) in 'Freedom and Resentment', which is not at the place at which Hieronymi suggests that we find a second charge of incoherence. In this further passage, concerned with the case of raising children and the attitude of the psychoanalyst to their patient, the argument is clearly for the *irrelevance* of determinism (see, Strawson, 2008a, p. 21).

The conclusion Strawson draws from this is *not* that the sceptic's suggestion is *incoherent* or *self-contradictory*. Rather, the conclusion Strawson draws is that determinism is *irrelevant*. Determinism is said to be irrelevant in the sense that it is not a consideration for which we actually do exempt agents; that it does not, actually, count as such according to our conceptual scheme and within our practice of responsibility. When Strawson, just a few sentences below, summarises the argument, that is just the thought he expresses:

when we do in fact adopt such an attitude [i.e., the objective attitude] in a particular case, our doing so is not the consequence of a theoretical conviction which might be expressed as 'Determinism in this case', but is a consequence of our abandoning, for different reasons in different cases, the ordinary inter-personal attitudes (Strawson, 2008a, p. 14)

And, as we've already seen, this thought is (as expected) repeated in section V, when summing up his argument in connection to the vicarious analogous of the personal reactive attitudes:

we must note, as before, that when the suspension of such an attitude or such [reactive] attitudes occurs in a particular case, it is *never* the consequence of the belief that the piece of behaviour in question was determined in a sense such that all behaviour *might be*, and, if determinism is true, all behaviour *is*, determined in that sense. For it is not a consequence of any general thesis of determinism which might be true that nobody knows what he's doing or that everybody's behaviour is unintelligible in terms of conscious purposes or that everybody lives in a world of delusion or that nobody has a moral sense, i.e. is susceptible of self-reactive attitudes, etc. In fact no such sense of 'determined' as would be required for a general thesis of determinism is ever relevant to our actual suspensions of moral reactive attitudes. (Strawson, 2008a, p. 19, original emphasis)

But that determinism is irrelevant *only* in the sense that it is not a consideration for which we actually do exempt agents—that it does not, as a matter of fact, count as such within our conceptual scheme—can't be the whole story, can it? Might it still not be that determinism *should* count as a reason to exempt? Isn't Strawson's argument that it *cannot* count as such?

The sense that this can't be the whole story seems to be what drives Hieronymi's interpretation. The driving force is Strawson's Promise—of the fundamental irrelevance of determinism for responsibility. This, I conjecture, is what leads her to make more of this point than just that we do not, now, actually, count determinism as a relevant consideration. Strawson's Promise is appealing precisely because it goes beyond the traditional quibbles and the

classic compatibilist strategy of clarifying a sense of, for example, ‘could have acted otherwise’ that is compatibilistic. That there is more to Strawson’s story has been and still is the general sense, and is part of the explanation for the great interest that the paper has undoubtedly attracted.

Benjamin De Mesel (2022c) has brought to our attention that with respect to just the argument that Hieronymi considers, Strawson admits of an inadequacy. The relevant exchange appears in *The Philosophy of P. F. Strawson* (1995), edited by Pranab Kumar Sen and Roop Rekha Verma, and published by the Indian Council of Philosophical Research (not to be mistaken for the later anthology with the exact same title, edited by Lewis Edwin Hahn (1998)). Rajendra Prasad’s (1995) paper in that anthology expertly deals with the argument for the irrelevance of determinism considered above. Importantly, the passage Prasad (1995, p. 357f.) reconstructs Strawson’s argument from is *exactly the same passage* as the one that Hieronymi (2020, p. 43f.) claims (implicitly) involves a second accusation of incoherence against the sceptic. Prasad’s charge against Strawson is that the argument is either inconclusive or question-begging. In his reply, Strawson straightforwardly pleads guilty as charged.⁶⁷ This is a striking interpretative finding. Not every detail of the exchange is of relevance here (cf., De Mesel, 2022c). What is illuminating for our discussion is what, at least in one dimension, the argument *was* and what Strawson’s defence is.

Prasad’s argument presents Strawson with a dilemma: either his argument is inconclusive or question-begging. We get the initial premise of the argument from Strawson’s survey of the reasons for which we actually excuse and exempt agents, those ‘special considerations’ in light of which we find it inappropriate to have reactive attitudes towards agents. Assume that we agree with Strawson that if any of the reasons that he lists are present, it is inappropriate to have reactive attitudes. Strawson then argues that it is not the case that if none of these considerations holds, that must be because determinism is not true. That is, to put it in its contrapositive, he argues that it is not the case that when determinism is true, this implies that some such special consideration is present. Thus, it is not the case that the truth of determinism would imply that no reactive attitudes are appropriate.

This is how Prasad (1995, p. 358f.) formalises Strawson’s argument (cf., De Mesel, 2022c, p. 3). Take D to be ‘Determinism is true’, $\neg R$ to be ‘It is

⁶⁷ Just a note on the chronology of ideas: Strawson visited India (for the third time) in the winter of 1987-1988, and a conference was arranged on his work by the Indian Council of Philosophical Research. I don’t know if Prasad was present or, for that matter, if he presented his paper to Strawson already at the conference in 1987. Perhaps.

inappropriate to feel or have any reactive attitudes' and [C] to be 'At least one consideration of the set of excuses and exemptions is present'. This, then, might be Strawson's argument for the irrelevance of determinism:

- (i) P1: [C] \rightarrow \neg R
- P2: \neg (D \rightarrow [C])
- C: \neg (D \rightarrow \neg R)

If that is Strawson's argument, however, then the argument appears *inconclusive*. As Prasad observes, there is no reason why determinism could not itself render reactive attitudes appropriate—i.e., make them inappropriate without therefore implying that [C].⁶⁸ Unless Strawson can show the *exhaustiveness* of some set like [C], then his argument will be inconclusive. If this presentation of Strawson's argument is to be valid—if it is to have the conclusion that *it is not the case that if determinism is true, then it is inappropriate to have reactive attitudes*—then Strawson's argument would need a further premise. The argument will be valid if we add the premise that, if it is inappropriate to have reactive attitudes, then at least one consideration of the set of excuses and exemptions is present. Amended, this would then be Strawson's argument:

- (ii) P1: [C] \rightarrow \neg R
- P2: \neg R \rightarrow [C]
- P3: \neg (D \rightarrow [C])
- C: \neg (D \rightarrow \neg R)

If (ii) is Strawson's argument, however, then the argument seems to be *question-begging*. As Prasad observes, the biconditional that the first two premises together form means that *nothing but* members of [C] make reactive attitudes inappropriate. If this is *assumed*, it is *assumed* that determinism cannot itself make reactive attitudes inappropriate. But that determinism does not, or cannot, make reactive attitudes inappropriate was exactly what the argument was meant to show. For this reason, the assumption (the biconditional) is question-begging.

⁶⁸ We could add to Prasad's point that, even if determinism does not imply any of the considerations in the set [C], and even if it does not itself directly render reactive attitudes inappropriate, it might nevertheless imply some other consideration, not yet taken into account, which is a member of some larger set [C]* that is the actual set of considerations for our practice.

Strawson's answer to Prasad's paper is quite sincere:

[Prasad] shows conclusively that one of my arguments against the relevance of the thesis of determinism to the question of moral assessment and responsibility is either invalid or question-begging. I argue that the truth of determinism would not entail the presence in any particular case of one or more of the normal 'inhibitors' of moral reactions. But do I contend that such presence is merely sufficient or also a necessary condition of the inhibition? If I contend only the former, the argument is inconclusive; if the latter, it begs the question. (Strawson, 1995, p. 430)

By the looks of it, Strawson *repudiates the argument*. This change of mind is undoubtably of great interpretative interest. As De Mesel remarks: "This is, as far as I can tell, the *only* place in Strawson's *oeuvre* where he repudiates an argument or claim in 'Freedom and Resentment'" (2022c, p. 4, original emphasis). (While De Mesel's subsequent discussion of Strawson's response itself bears great interest, it is not the topic of this chapter. We'll have occasion to revisit the exchange in the next chapter.)

The purpose of the preceding point is twofold. First, to show that Strawson's ambitions indeed were great. That Strawson himself came to see that his ambitions—at least, in a particular instance, with respect to "*one of*" his arguments for the irrelevance of determinism—were *too* great reveals that his ambitions in 'Freedom and Resentment' *actually were great*. In this respect, Hieronymi is right (as others also are) to try to make sense of Strawson's approach (in 'Freedom and Resentment', at least) such that (something like) Strawson's Promise is an intelligible part of that approach.⁶⁹ This fact buttresses an assumption of this thesis concerning the dialectic though which we may fruitfully understand Strawson's approach.

The second purpose of the preceding point, pertaining directly to the issues of this chapter, is to show that Strawson's argument for the irrelevance of determinism indeed is (a) for the *irrelevance* of *determinism* and not, *pace* Hieronymi, an argument (b) for the *incoherence* of *scepticism*.⁷⁰ On Prasad's presentation of Strawson's argument, it is for (a) and not (b). The argument

⁶⁹ De Mesel (2022c, p. 7) argues that Strawson "knew that his account in 'Freedom and Resentment' was vulnerable to the kind of criticism formulated by Prasad" since he develops his view, in just those respect which would address the concern Prasad raises, in "Liberty and Necessity" (2011c) and "Freedom and Necessity" (1992b). (Indeed, he develops it in just those respects in which he develops his view in direct response to Prasad's criticism). In this connection, it is perhaps noteworthy that "Freedom and Necessity" was first published already in 1985 in the original, French version of *Analysis and Metaphysics*.

⁷⁰ For another clear expression of this fact, see Alvarez (2021, esp. p. 192).

that Prasad indicts is, Strawson (1995, p. 430, emphasis added) himself says, “*one of my arguments against the relevance of the thesis of determinism*”.⁷¹ Prasad, by Strawson’s own lights, correctly represents Strawson’s argument in the relevant passage. Strawson’s response to Prasad is not, then, of any help for the Statistical Reading with respect to the puzzles above; if anything, to the contrary.⁷²

If you allow me to hypothesise why Strawson might be thinking that determinism would not *itself* make any reactive attitude inappropriate and, relatedly, why Hieronymi has gotten the impression that the fact that determinism is a *general thesis* should be a central fact to the argument by Strawson here reviewed, then I would hypothesise that the reason is that Strawson excludes the ‘general thesis’ of determinism as itself relevant because he eschews the existence of a ‘general reason’ for why we have these practices. If there is a “conflation” of two notions in ‘Freedom and Resentment’, I suspect that it is this one: between ‘general reason’ and ‘general thesis’. How could it be that determinism *itself* would show that any reactive attitude would be inappropriate? I don’t see it. It is more plausible that determinism would *not*, strictly speaking, *all by itself* do this. The closest we come to this thought is, rather, that there would be *some very general condition* on which the appropriateness of *any* reactive attitude depended which would be undermined by the truth of determinism. Prasad’s suggestion is no more than a logical possibility; we have been given no reason to think it is actually so, neither any assistance in discerning the conceptual connections that are necessary if we are to make sense of how determinism itself could exempt everybody, that this inference would be an actual feature of our conceptual scheme. (If we tried to find a name for that peculiar, very general condition, perhaps we could find no better candidate than “free will”).

The assumption that there would be a ‘general reason’ of this kind is alien to Strawson’s mind. On Strawson’s picture, “when we do in fact adopt [the objective attitude] in a particular case, our doing so [...] is a consequence of

⁷¹ Consider Strawson’s formulation: “*one of my arguments against the relevance of the thesis of determinism*”. This suggests that this is *not the only argument* to that conclusion of Strawson’s. If that is right, it may be too quick to say, as Strawson himself suggests we perhaps want to say, that he has “dwindled into a mere compatibilist” (1995, p. 431). I thank Maria Alvarez for prompting me to pay additional attention to this passage.

⁷² That is, it undermines the claim that the relevant passage embodies a charge of incoherence. This is “*if anything, to the contrary*” because that the passage does not embody such a charge would only really be relevant for the merits of the Statistical Reading if it was not the case that the Second Incoherence Puzzle would still not be answered by finding a second charge of incoherence in *this* passage.

our abandoning, *for different reasons in different cases*, the ordinary interpersonal attitudes” (Strawson, 2008a, p. 14, emphasis added). It is not because we have some ‘general reason’ to do this: “It needs no saying how multifarious these considerations are”—i.e., how various and diverse our reasons for excusing and exempting are (Strawson, 2008a, p. 7). Our reasons for excusing and exempting are context-specific and reflect that range of distinctions that we are all too prone to abstract away from, “especially in our cool, contemporary style” (Strawson, 2008a, p. 7). One reason why we struggle so to get a clear view of our concepts and practices, Strawson think, is because of “the characteristically philosophical compulsion to seek and find the roots of the institution [of morality] in some single, unitary source, be it reason, God, human emotions, social utility or some other philosophical darling” (1995, p. 433). A failure to see how determinism could by itself otherwise—i.e., without in fact presupposing that there is some ‘general reason’ that it triggers—so smooth out our world that in it there were no such distinctions, that failure, so I conjecture, might be the explanation why Strawson takes this argument to rule out determinism as relevant.

V. The Self-Contradiction Puzzle

Returning to the issue of a second accusation of incoherence. It might be that, even if the passage which Hieronymi takes to embody the second charge of incoherence in fact manifests an argument for the irrelevance of determinism (as argued in the last section), that passage may nevertheless *also* embody a charge of incoherence. Conceding this, let’s next consider Hieronymi’s reconstruction of what she takes to be at least one argument of Strawson’s. This lends occasion for us to bring to the fore a central feature of the Statistical Reading that is, in light of previously unpublished material, uneasily attributed to Strawson.

According to Hieronymi, the survey of our reasons for exemption in the relevant passage is part of Strawson’s attempt to entrap the sceptic in a contradiction. Here is Hieronymi’s (2020, p. 43) reconstruction of the argument of the relevant passage:

- P0: When we exempt a person, the reason for which we adopt the objective attitude is that we see the agent as incapacitated in some or all respects for ordinary interpersonal relationships.

- P1: If determinism is true, then all behaviour is determined* (i.e., determined in the sense that all behaviour is if determinism is true).
- P2: Determinism might be true.
- P3: When we exempt a person, the reason for which we adopt the objective attitude is that we see the agent as determined*.

The self-contradiction would lie in affirming all of P0-P3. In particular, we cannot hold both P0 and P3, because, on the Statistical Reading, the capacities for ordinary interpersonal relationships are (partly) determined by the *statistically* normal capacities of a given community, so nothing statistically normal could be a reason to exempt—and, given P1, being determined* is as statistically normal as anything can be.

Notably, the supposed accusation of incoherence or self-contradiction is far from explicit in the text. The contradiction rather emerges first *if* we understand Strawson as on the Statistical Reading. But if so, its role in motivating that reading to begin with is questionable. However (to once more borrow a phrase from Strawson), this dismissal might seem altogether too facile; and so, in a sense, it is. I suppose Hieronymi takes the above argument not to presuppose but to *evince* Strawson's understanding of the workings of excuses and exemptions, as presented by the Statistical Reading. Let's grant this. It is the *conclusion*, not the *argument*, that requires, interpretatively, that we assume (something like) the Statistical Reading, and thereby that we ascribe that reading to Strawson. And it is then that we may here note yet another instance of the charge of incoherence or self-contradiction.

As Hieronymi reconstructs the argument, a crucial premise that the sceptic and all other parties must concede is P1—that is, “that if there is a coherent thesis of determinism, then there must be a sense of ‘determined’ such that, if that thesis is true, then all behaviour whatever is determined in that sense” (Strawson, 2008a, p. 11). This fact about determinism—that it is “a general thesis” (Hieronymi, 2020, pp. 11, 14, 18, 19, 23, *passim*), meaning that it is “true of everyone at all times” (Hieronymi, 2020, p. 1)—is what makes it illegitimate as, even inconceivable as, even logically excluded from being, an excusing or exempting condition.⁷³ If we assume the Statistical Reading, this

⁷³ “Strawson means to use the fact that determinism is a *general* thesis, true of everyone at all times, together with the account he has given of the ways in which the reactive attitudes are and should be modified or suspended, to reach the conclusion that acceptance of the truth of determinism neither would nor should lead to the decay or repudiation of the reactive attitudes” (Hieronymi, 2020, p. 16, original emphasis).

follows. But to so understand P1 is not an unproblematic interpretative assumption.

Consider Russell's (2017a, 2017b, 2017d, 2021) objection to "the central argument": *the Capacity Objection*. According to Russell, the charge of incoherence is due "to [Strawson's] conflation of abnormality and incapacity" (2021, p. 759). "The central argument" fails because the reason for which we exempt someone is not that they are "abnormal"; it is that they do not have the relevant capacities. And, at least by Russell's lights, "it is not inconceivable or self-contradictory to suggest that there could be a world, or things might develop, such that everyone is or becomes *incapacitated*" (2017d, p. 43). Call this *the possibility of universal incapacitation*. Here is an outline of how the line of reasoning for this possibility goes: if X is a capacity necessary for being a responsible agent, and we know that some individuals are not responsible agents in virtue of the fact that they lack X , would it not be possible for anyone that they would, like some presently do, lack X ? If this is true about anyone, then why could it not possibly be true of everyone that they lack X ?⁷⁴ Whether it is in fact true of everyone, as whether it is true of someone, will of course depend also on the further fact of whether the relevant agents in question have the requisite capacity. What Hieronymi calls "the central argument" is "too facile" on Russell's reading, representing as it does the Conventional Way. "The central argument", as it stands—i.e., without correcting for the supposed equivocation of 'abnormal' and 'incapacity'—would have to deny the possibility of universal incapacitation. According to the Capacity Objection, this denial is implausible.

According to Hieronymi, an argument like the Capacity Objection "depends on the claim that there are moral standards that all ordinary interpersonal relating could fail to meet, requiring us to adopt the objective attitude universally" (Hieronymi, 2020, p. 71f.).⁷⁵ On the Statistical Reading, "Strawson's rejects such standards" (Hieronymi, 2020, p. 72). He "rule[s] out the possibility that anything true of everyone is, in fact, a reason to exempt"

⁷⁴ Note that, if sound, the elenctic transcendental argument would block the move from 'anyone' to 'everyone'. The latter amounts to the rejection of the scheme within which we make sense of excusing or exempting anyone.

⁷⁵ Hieronymi says this about 'the generalization strategy'. As the argument is presented above, it includes the relevant features of the generalization strategy in Russell's 'capacity objection', preserving what is necessary for the applicability of Hieronymi's response. A note on Hieronymi's response: perhaps she should say that the capacity objection depends on the claim there *could be* moral standards, rather than that there *are* moral standards, that all ordinary interpersonal relation could fail to meet. A *sceptical* conclusion from the capacity objection depends on that there *are* such standards.

(Hieronymi, 2020, p. 99, also pp. 18, 33, 37, 51, 101). The reason for this, on the Statistical Reading, is that it is a condition on the possibility of a human society that certain demands and expectations are generally met, so, given that we live in a human society, the demands and expectations will have to be such that *most* have the capacity to meet them (and indeed in fact generally meet them), and being able to meet these demands and expectations just is what it is to be a responsible agent; thus, *most* will be responsible agents—“we can rest assured that nothing true of everyone will provide a reason to exempt” (Hieronymi, 2020, p. 106).⁷⁶ There are not, and could not be, on this interpretation of Strawson’s argument, any “standards on ordinary relating that are universally unmet” (Hieronymi, 2020, p. 101).

The Statistical Reading, thus explicated, faces a further interpretative puzzle—call it *the Self-Contradiction Puzzle*. In a previously unpublished letter addressed to Paul Russell (see Appendix), responding to his “Strawson’s Way of Naturalizing Responsibility” (2017d), Strawson writes:

⁷⁶ It is perhaps noteworthy that there is a discrepancy on Hieronymi’s presentation of the argument between what are essential commitments of Strawson’s and what is necessary for ‘the central argument’ on the Statistical Reading. Is Strawson’s ‘core idea’ really that “nothing true of *everyone* could give us reason to suspend the reactive attitudes” (Hieronymi, 2020, p. 18, emphasis added), even on the Statistical Reading? If only conditions representing a statistical abnormality exempt, then while it would be *sufficient* if everyone fulfilled some condition for us to know that this condition is not an exempting condition, it would not be *necessary* that some condition was the universal condition, “true of everyone”, for this conclusion. It would be enough if *most* (statistically speaking) fulfilled that condition for that condition to not be an exempting condition, on the Statistical Reading. The slogan should be: ‘nothing true of *most* could give us reason to suspend the reactive attitudes’. In this respect, Strawson’s argument on the Statistical Reading is even more clearly a more demanding argument than what an elenctic transcendental argument would be, as the latter only charges scepticism with incoherence if scepticism is the denial of the *entire* framework. This, furthermore, might suggest that the Statistical Reading devours more than is desirable, as it excludes a quite conceivable (compatibilistic) possibility. (However, Hieronymi is not perfectly consistent in her presentation of the Statistical Reading in a, for this criticism, relevant respect; so, it might be that this criticism misrepresents the position. Sometimes, that is, the claim is not as above but is rather that “the fact of our natural human commitment to characteristically interpersonal relating pushes us to adjust our standards towards the majority *or the dominant*” (2020, p. 90, also pp. 84, 91, 95). That is, what is *statistically* normal is not (the only) fact that defines the standards, but (rather or also) that of being ‘normal’ according to the prevailing ideology of the society—we might say, *not* being a minority in the sense that women might be a minority even if they (statistically speaking) represent (roughly) half of the population. If that is indeed the (or a) relevant sense, it is unclear how the criticism raised here applies.)

I do not want to rule out the idea of universal incapacitation as incoherent or logically impossible.

This spells trouble for the plausibility of the Statistical Reading as an interpretation of Strawson's approach. The Statistical Reading treats the exclusion of the possibility of universal incapacitation as central to Strawson's approach, as a key claim of "the central argument", anchoring the self-contradictory nature of the sceptical alternative—but then Strawson himself admits of it as (at least) a logical possibility; as *not* an incoherent idea. On the Statistical Reading, it must seem like Strawson is here contradicting himself regarding the possibility of universal incapacitation.⁷⁷ This is puzzling; this is our further interpretative puzzle:

The Self-Contradiction Puzzle: If Strawson's "central argument" excludes the possibility of universal incapacitation, then why does Strawson explicitly admit of the possibility of universal incapacitation as not incoherent or logically impossible?

How might a proponent of the Statistical Reading deal with this? The letter is dated August 9th 1992. Indeed, Strawson might have changed his mind in those forty years between publishing the paper and writing the letter. A change of mind on the issue might resolve the Self-Contradiction Puzzle. However, would it not be strange if Strawson, especially when directly responding to a paper on his view that deals with "the central argument", did not then indicate such a change of mind?⁷⁸

What Strawson says in the rest of that paragraph of the letter is also instructive given our present concerns. Here is the paragraph in full length:

I do not want to rule out the idea of universal incapacitation as incoherent or logically impossible. But the thoroughgoing Pessimist must hold that the truth of determinism would have universal incapacitation as a consequence. I do not believe that and neither, I think, do you. (Strawson, 1992c)

⁷⁷ It "must" on the assumption that Hieronymi's reading unambiguously affirms that the relevant sense of 'normalcy' that determines the conditions for exemption is the statistical sense. But, as noted in the last footnote, Hieronymi is not perfectly consistent on what the relevant sense of 'normalcy' is.

⁷⁸ Would it not be only decent to let Russell know—if that was indeed the case—that Russell has correctly understood the argument of the paper, that his criticism is valid, but that it does not apply to Strawson's view anymore since there has now been a change of mind?

This might at first seem like it provides an opening for the Statistical Reading. Might Hieronymi not claim that it is precisely because the relevant universal incapacitation is that which would follow from *determinism* in particular, and that it is because determinism is a general thesis that it cannot be that “the truth of determinism would have universal incapacitation as a consequence”? Then it would be possible that universal incapacitation was the conceivable result of some *other* consideration being constantly present (for example, if a fungus mutation that takes the human brain as its host were to become pandemic)? But this would not be enough. On the Statistical Reading, in Hieronymi’s words, “Strawson thinks that *nothing* true of everyone could provide an exemption” (Hieronymi, 2020, p. 33, emphasis added). Perhaps the possibility of universal incapacitation is only a possibility as a result of *multiple* reasons, some true of some, others true of others, such that they are only jointly sufficient for universal incapacitation? I leave this possibility for the Statistical Reading to address.⁷⁹

The Self-Contradiction Puzzle is not only puzzling on the Statistical Reading. On the Conventional Way, “the central argument” is said to be an argument for incoherence but is considered a bad argument—e.g., since equivocating between ‘abnormality’ and ‘incapacity’. The standard motivation for finding it a bad argument, and thus for the Conventional Way with the Incoherence Puzzle, is the Capacity Objection. However, the Capacity Objection, just as the Statistical Reading, assumes that Strawson’s argument for incoherence, “the central argument”, is meant to exclude the possibility of universal incapacitation—that is why the possibility is held against him; why the argument is bad. And yet, Strawson straightforwardly admits of this possibility. Resultantly, also the Conventional Way faces a version of the Self-Contradiction Puzzle. Because also the Conventional Way charges Strawson with a self-contradiction. Despite the explicit admittance of the possibility of universal incapacitation, they both attribute to him the rejection of that possibility (as being incoherent or logically impossible).

What Strawson’s letter indicates is that “the central argument” is not for the incoherence of universal incapacitation. In his response, what he stresses is not the incoherence of scepticism but, immediately after admitting of the logical

⁷⁹ We may ask if Strawson’s admittance of the possibility of universal incapacitation does not, just as for the Statistical Reading, spell trouble for a reading that attributes to him an elenctic transcendental argument against the responsibility sceptic? It does not, not in the same way. Remember that, on an elenctic transcendental argument, it is *the challenge* that is incoherent. It is not the *possibility* that the sceptic takes their argument to show to be actual that is incoherent or self-contradictory. It is rather *the argument* for that conclusion that is purported to be incoherent.

possibility of universal incapacitation, what he stresses is the *irrelevance of determinism* for whether anyone or everyone is incapacitated.

VI. Conclusion

In light of this, what are we to say about “the central argument”? I’m not sure. It should plausibly not be rendered in any way that ascribes so much as the attempt to show the incoherence or logical impossibility in universal incapacitation.

A *possible* reading, close to hand in light of what we’ve seen, is that the passage—ending with, “But it cannot be a consequence of any thesis which is not itself self-contradictory that abnormality is the universal condition (2008a, p. 12)—expresses a charge of incoherence against the responsibility sceptic nevertheless, albeit in *the form of an elenctic transcendental argument*. The elenctic transcendental argument would then be, very roughly stated, that taking the set of considerations for which we exempt as given by Strawson’s analysis, and assuming that each consideration is actually an exception of sorts—expressive not of the *default*, but of an *abnormality* in interpersonal relationships—and because the sceptic must argue that one of these holds universally in virtue of the truth of determinism, the sceptical conclusion would have us lose touch with the conditions for when we exempt—for an exception cannot be the rule. The claim that a consideration holds universally does not present an unintelligible, incoherent or logically impossible scenario. In forcing us to lose touch with the considerations, so the argument would go, the sceptic’s argument undermines itself, because our grip on our considerations for exemption was supposed to provide the ground for the sceptical conclusion. Such an anti-sceptical argument does not imply that universal incapacitation is incoherent or logically impossible, but it explains the charge of incoherence. The charge of incoherence consists, on this explanation, in rejecting our conceptual scheme for reasons that only count as such reasons if we presuppose that very conceptual scheme. These cannot be our reasons for rejecting it unless the conceptual scheme is not, actually, to be rejected; if it is to be rejected, then these considerations do not count as the argument presupposes. This, obviously, shares some crucial features of the Statistical Reading, but does not share the latter’s exclusion of the possibility of universal incapacitation—thus, unlike the Statistical Reading and the Conventional Way, it avoids the Self-Contradiction Puzzle.

In an earlier draft, I entertained the idea that, if we're set on taking the Incoherence Puzzle seriously—an assumption we evidently have reason to question—then, we should adopt this reading, with a crucial qualification—viz., that Strawson does not himself think that it quite works. The qualification is necessary if this reading is to handle the Second Incoherence Puzzle. What we've called “the central argument” is, on this reading, indicative of the *ambition* of an elenctic transcendental argument against the responsibility sceptic. What we get, we should understand as an attempt; or, more charitably, an ambition; or, less charitably, as a lapse, a giving into temptation, an exaggeration. The important point is that it is not the central argument of ‘Freedom and Resentment’, not by Strawson’s own lights anyway; this was meant to explain why it does not reappear. I’ve lost confidence in this reading: it does not strike me as either very plausible or interesting interpretation anymore. (I might be wrong; it would not be the first time.)

We have not fully considered the interpretative merits of Hieronymi’s novel reading of Strawson’s approach. What has been shown, however, is that a central motivation for considering the reading as a whole is lost on closer inspection. The claim that “the central argument” is Strawson’s *central* argument generates a very surprising disanalogy—the key claim of “the central argument” does not reappear, even though the overarching argument is undoubtably an argument by analogy, and even though all the other central tenets of that overarching argument do reappear. This is the Second Incoherence Puzzle. And, because the Statistical Reading commits Strawson to the claim that the possibility of universal incapacitation is incoherent, the Statistical Reading also faces the Self-Contradiction Puzzle. While we have not properly considered all the interpretative merits of Hieronymi’s novel reading, it is hard to see how it may overcome these issues.

If we take the Incoherence Puzzle seriously, as Hieronymi asks us to do, this does not motivate the Statistical Reading—it only ensnares us in further puzzles, leaving us in even greater puzzlement. Returning to “the central argument” now, we at least recognise, not only that the Statistical Reading misunderstands the argument but, furthermore, that the Conventional Way also misunderstands it—for the very straightforward reason that its understanding of the argument is fundamentally the same. The Statistical Reading and the Conventional Way do not differ in what they take “the central argument” to be, but only with respect to what *plausibility* and *centrality* they take it to have. The former takes Strawson’s endorsement of optimism at the end of ‘Freedom

and Resentment’ to pivot on “the central argument”, and takes that argument to be “a powerful argument” (Hieronymi, 2020, p. 16, also, pp. 2, 71). The latter finds the argument implausible—it finds it “misleading and mistaken to place any emphasis on considerations of ‘abnormality’” (Russell, 2017d, p. 43)—but leaves it open that much else besides “the central argument” may be of interest for the optimistic project.

Because they do share the same understanding of the argument, however, the reason that the Conventional Way does not, unlike the Statistical Reading, find the argument plausible, must be said to be confused. The Self-Contradiction Puzzle shows as much.

This does not mean that we should not, just as the Conventional Way does, not take the Incoherence Puzzle seriously. A new reason for this has emerged in the course of this chapter: if we do take the Incoherence Puzzle seriously, this lands us in confusion; by all appearances, then, we stand a better chance of making sense of Strawson’s approach if we do not take it seriously.

A further reason not to take the Incoherence Puzzle seriously may be derived from a kind of generalisation of the Second Incoherence Puzzle. It is not only a problem, as per the Second Incoherence Puzzle, that “the central argument” does not reappear in section V, seeing that it is supposed to be central and given the analogical structure of ‘Freedom and Resentment’. The claim that the argument is to be central also face the issue of explaining, not just why it does not reappear a second time, but why it does not *ever* reappear. Nowhere does the claim “abnormality cannot be the universal condition” or an equivalent of this reappear in what more Strawson did write on responsibility (cf., Strawson, 1980, 1992a, 1995, 1998b, 1998e, 2008c, 2008d, 2011c; Strawson et al., 2008).

Perhaps we have discovered that, if we are to unearth an argument for the irrelevance of determinism, we do best to consider as irrelevant “the central argument”. With respect to “the central argument”, we have only found new reasons for the conventional response. We have learned that we have no other alternative, or no coherent alternative, for the moment, than to disregard it.

4. The Reversal Move

I. Introduction

The Reversal Move proposes that Strawson reversed the order of explanation assumed by the traditional approach. According to the Reversal Move, we should understand what it means to be responsible as (in some sense) *constituted, determined or fixed by* our practices of holding responsible and having reactive attitudes, rather than the other way around. For the Reversal Move, this is the canonical passage in ‘Freedom and Resentment’:

Only by attending to this range of attitudes [i.e., the reactive attitudes] can we recover from the facts as we know them a sense of what we mean, i.e. of *all* we mean, when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation, and justice. But we *do* recover it from the facts as we know them. We do not have to go beyond them. (Strawson, 2008a, p. 24, original emphasis)⁸⁰

The Reversal Move certainly has that revolutionising air about it. It is the most salient approach for anyone seeking to make good on Strawson’s Promise: the sense that something about the very way in which Strawson approaches the issue of free will and responsibility shows that the truth of determinism is irrelevant. This is the thought that the Reversal Move elaborates on. As it is typically presented, the Reversal Move is supposed to be anti-sceptical or, at the very least, to not be beholden to the traditional worries about freedom of the will. A failure in this respect does not only count against a proposed account of the Reversal Move but raises the question whether the account could even be said to be an account of the Reversal Move.

Beyond these schematic statements, however, it is rather difficult to say what this radical new approach is and what it is supposed to be. The range of

⁸⁰ The passage’s canonical status is evident from the fact that it is cited as expressive of the Reversal Move by, for example, Beglin (2018, p. 615; 2020, p. 2243), De Mesel (2022a, p. 1894), De Mesel & Heyndels (2019, p. 799), and Shoemaker (2017, p. 481).

views that all have some claim to being articulations of the Reversal Move is quite diverse (e.g., Beglin, 2018; Beglin, 2020; Bengtson, 2019; Bennett, 1980; De Mesel, 2022a; De Mesel & Heyndels, 2019; D. Shoemaker, 2017, 2022; Wallace, 1994; Watson, 2008, 2014).⁸¹ They operate at different *levels* of explanation, with different *kinds* of explanations, with different ancillary *interpretative assumptions*, and with different *ambitions*.⁸² Patrick Todd's (2016) review of some of the ways in which the Reversal Move has been presented reflects this difficulty; his argument is in effect an *argumentum ad ignorantiam*.⁸³ His confusion is not unwarranted: it is far from clear how exactly the Reversal Move is to be stated, especially if it is supposed to deliver on what at least seems to be its promise, of transcending the traditional debate and vindicating optimism about responsibility.

When we say that Strawson proposed a reversal of the order of explanation, we need to ask, among other things, what is the traditional approach, the order of which is reversed? What kind of explanation is at issue, in the traditional approach and in Strawson's approach? If there is a kind of explanation with respect to which the order on Strawson's approach and on the traditional approach may be said to be inversely related, how much of Strawson's approach may be captured in terms of such a move and, especially, how plausible is the claim that this move itself may make good on Strawson's Promise?

Interpretatively, it is relevant to note, already from the start, that the Reversal Move is not explicitly presented (and surely not explicitly presented *as such*) by Strawson.⁸⁴ It has emerged from the secondary literature, and it is the

⁸¹ For some further presentations of the claim, see, for instance, Brink & Nelkin (2013), Clarke & Piers (2023), Coates & Tognazzini (2013), D'Arms & Jacobson (2022) McKenna (2012, ch. 2), and Tognazzini (2013).

⁸² Wallace (1994), for example, intentionally construes the Reversal Move such that it itself will not rule out incompatibilism, and it is a central question for the proponents of the Reversal Move whether he does not, then, really articulate any version of the Reversal Move.

⁸³ It is only *in effect* that, and that qualification is enough for it not to be a logical fallacy. More charitably, the argument could be presented as a burden of proof dumping argument: it is for the Reversal Theorists to explain how the Reversal Move is supposed to work and how much it may achieve. By Todd's own account, his argument is rather one pivoting on a dilemma: either the Reversal Move is not antilibertarian (as its proponents suggest that it is) or it is simply implausible (Todd, 2016, p. 223).

⁸⁴ A formulation by Beglin suggests that the passage quoted above itself supplies "something close to an articulation of the Strawsonian reversal" (2018, p. 616); this, I think, is a bit of an exaggeration, revealing in that Strawson may, at best, be said to supply *something close* to a presentation of his own reversal.

secondary literature that we will be preoccupied with in this chapter. Partly because the Reversal Move has been developed without much concern for the interpretative issue, it is difficult to get a grasp of what the move is even supposed to be. This is not just the problem of saying *what* the Reversal Move is. It remains a real question whether there even is *anything* like it on Strawson's approach (Alvarez, 2021, p. 194).

This chapter considers what the Reversal Move is and what it has to be if it is to make good on Strawson's Promise. We will begin (in section II) with a few remarks on the history of the Reversal Move, primarily to retrieve a sense of what it is (or, what it was) supposed to be. Then (in section III) we'll consider two accounts of the Reversal Move. The first is David Shoemaker's (2017, 2022) *Fitting Response-Dependence Thesis*. The second is Benjamin De Mesel's (2022a) (and Sybren Heyndel's (2019)) *Meaning-Based Account*. We then turn (in section IV) to assess their chances of making good on Strawson's Promise. While the Meaning-Based Account has a better claim to capturing Strawson's view, we might begin to worry whether the revolutionary air that accompanied the Reversal Move has not disappeared and, if so, whether the account is to count as an expression of the Reversal Move. The Fitting Response-Dependence Thesis may have a better claim to expressing the Reversal Move, but this depends on how it is elaborated.

In light of this discussion, I conclude that what is sought on the Reversal Move, or what is needed for it to work, is a *naturalistic explanation* that shows why the concept of responsibility that we have is one for which the truth of determinism is irrelevant. Attention to how we react to other people's behaviour and our own, rather than to some independent or prior nature of the concept of responsibility itself, seems to be the way forward. If we are to vindicate that elusive sense that we may circumvent the traditional problem, the Meaning-Based Account would have to be supplemented with such a naturalistic explanation and the Fitting Response-Dependence Thesis would have to develop its naturalistic explanation further. This sets the stage for the following chapters.

II. The History of the Reversal Move

By Way of Introduction

The claim that Strawson presents something like the Reversal Move is sometimes (e.g., Todd, 2016) traced to Gary Watson's (2008) reading in

“Responsibility and the Limits of Evil: Variations on a Strawsonian theme”. But we should probably trace the attribution of such a move a bit further back, to Jonathan Bennett’s (1980) “Accountability”.⁸⁵ Bennett writes of Strawson’s approach:

feelings are made central, and are not tied systematically to any propositions about their objects. My feeling of indignation at what you have done is not a perception of your objective blameworthiness [...]. It expresses my emotional make-up, rather than reflecting my ability to recognize a blame-meriting person when I see one. (Bennett, 1980, p. 24)

We have here more than one characteristic feature of the Reversal Move:

- (I) A contrast with the *traditional* approach: “My feelings of indignation at what you have done is *not* a perception of your objective blameworthiness”.
- (II) An abandonment of the idea of an *independent* notion of responsibility or *independent* facts of responsibility: indignation is *not* “a perception of your *objective* blameworthiness” (cf. e.g., Coates & Tognazzini, 2013; McKenna, 2012; D. Shoemaker, 2017; Tognazzini, 2013; Wallace, 1994; Watson, 2004, 2014).⁸⁶

Another element of Bennett’s reading, however, has a more complicated relation to the Reversal Move: Bennett is explicit that he reads Strawson as taking the reactive attitudes to be “non-propositional” (1980, p. 24).

- (III) Non-propositionalism: the reactive attitudes “are not tied systematically to any proposition about their objects”.

Bennett’s attribution of non-propositionalism is grounded in another idea, also hailed as characteristically Strawsonian—what has come to be known as an Affective Account of blame (or, more generally, of what it is to hold someone

⁸⁵ Not to be confused with his “Accountability (II)” (2008), written specifically for McKenna and Russell’s *Free Will and Reactive Attitudes* (2008), and though largely overlapping, is not identical in content.

⁸⁶ As Bennett puts it in “Accountability (II)” (2008, p. 55), “I think that many people have a notion of accountability which incorporates the belief that desert or blameworthiness of accountability is strictly a matter of objective fact” and, in this respect, he takes Strawson’s approach to be “more revisionary—or rather, excisionary” than Strawson’s presentation may lead us to think.

responsible) (cf., Carlsson, forthcoming; Tognazzini, 2013).⁸⁷ On the Affective Account, holding responsible is understood in terms of the reactive attitudes, understood as a particular set of *emotions* or *sentiments*. While the Reversal Move and the Affective Account are usually developed together or in tandem (cf., McKenna, 2012; D. Shoemaker, 2017, 2022; Wallace, 1994), the Reversal Move and the Affective Account are distinct, neither entailing the other, even if it might be that one encourages the other.⁸⁸ The Affective Account is an inessential part of the Reversal Move; an articulation of that move is not disqualified as an articulation of that move simply because it does not subscribe to the Affective Account.

Compare the passage from Bennett (1980, p. 24; also, 2008, p. 55) with Watson's, possibly canonical, articulation of the Reversal Move:⁸⁹

What these otherwise very different views [the consequentialist-optimist and the pessimist] share is the assumption that our reactive attitudes commit us to the truth of some independently apprehensible proposition which gives the content of the belief in responsibility; and so either the search is on for the formulation of this proposition, or we must rest content with an intuition of its content. [...] ⁹⁰

In Strawson's view, there is no such independent notion of responsibility that explains the propriety of the reactive attitudes. The explanatory priority is the other way around: It is not that we hold people responsible because they are responsible; rather, the idea (*our* idea) that we are responsible is to be

⁸⁷ The interpretative issue of whether Strawson does in fact himself hold an Affective Account of blame (or responsibility generally) need not detain us here.

⁸⁸ This is not a controversial claim, but only a reminder. Michael McKenna's presentation, for example, implies the same view: "I develop P. F. Strawson's theory of moral responsibility and endorse *two* crucial elements of it. One is that being morally responsible must be understood by reference to the nature of holding morally responsible. Another is that holding morally responsible ought to be understood by reference to a particular range of moral emotions and their related practices" (2012, p. 31, emphasis added). For another example, see Tognazzini (2013, p. 1300).

⁸⁹ Pace Todd (2016, p. 210, n. 4), who takes Bennett to *not* attribute a Reversal Move to Strawson, albeit, in pointing to (the equivalent of) the passage cited above, he does note that this might not be quite right. Shoemaker (2017, 2022) rightly credits Bennett with attributing a Reversal Move to Strawson.

⁹⁰ The passage continues: "For the social-regulation theorist [i.e., the consequentialist-optimist], this is a proposition about the standard effects of having and expressing reactive attitudes. For the libertarian, it is a proposition concerning metaphysical freedom. Since the truth of the former is consistent with the thesis of determinism, the consequentialist is a compatibilist; since the truth of the latter is shown or seen not to be, the libertarian is an incompatibilist."

understood by the practice, which itself is not a matter of holding some propositions to be true but of expressing our concerns and demands about our treatment of one another. (Watson, 2008, p. 117, original emphasis)

As in the passage from Bennett above, we find in Watson's formulation two central characteristics of the Reversal Move: (I), the *contrast* with the *traditional* view and, in making this contrast, (II), a *rejection* of an "*independent* notion of responsibility". Indeed, we also find (III), that the practice is "not a matter of holding some propositions to be true but of expressing our concerns and demands about our treatment of one another".

As said, the non-propositionalism has a complicated relation to the Reversal Move. This is so, for one thing, because (III) is understood as involving a kind of *non-cognitivism* and, for that reason, rejected by some who nevertheless take themselves to be developing something like the Reversal Move (e.g., McKenna, 2012; Scanlon, 2003; Wallace, 1994).⁹¹ However, the non-propositionalism is not necessarily rejected *as a reading* of Strawson, which leaves the association of Strawson's approach and non-cognitivism open.⁹² At least by these authors, (III) is treated as an inessential aspect of the Reversal

⁹¹ (III) has also been cause for straightforwardly rejecting the Reversal Move. For example, in distancing herself from Strawson's approach, Nomy Arpaly (2006, p. 28) writes, "contrary to those who think that blameworthiness and praiseworthiness of the agent for her action are epiphenomena of the reactive attitudes, it is the blameworthiness or praiseworthiness of the agent for her action that makes her an 'appropriate' object of the emotions, and not the other way around. If we simply follow the moral emotions in order to figure out who is blameworthy and who is not, we are bound to be misguided". (I owe the association of Arpaly's statement with criticism of the Reversal Move to Audun Bengtson (2020, p. 28).)

⁹² It is a fact that Strawson (1980) does not, in his response to Bennett, repudiate the non-propositionalism claim on Bennett's reading. We might wish to add that it is, equally, a fact that he does not explicitly endorse it. However, he does say that "Bennett [...] sets out and elaborates the essence of my position with such thorough and sympathetic understanding as to leave me little to say beyond recording my admiring appreciation (Strawson, 1980, p. 264). Shoemaker (2017, p. 494) takes Strawson to have a non-cognitivist understanding (at least) of a (or some) basic reactive attitude(s), (e.g.) resentment (see also, Deigh, 2011). D'Arms & Jacobson (2022) argue that, to avoid certain crucial challenges, the Reversal Move must be developed in terms of 'natural emotions', i.e., not 'cognitively-sharpened' emotions. Add to these considerations, also, that Strawson's review of "the particular conditions in which [the reactive attitudes] do or do not seem natural or reasonable or appropriate" explicitly concerns "what sorts of special considerations might be expected to modify or mollify" a reactive attitude (2008a, p. 7). Strawson does not seek to show the irrelevance of determinism by showing the irrelevance of *any* considerations about the appropriateness of the reactive attitudes (Alvarez, 2021). And, as we had occasion to consider in the last chapter, one of his arguments for the irrelevance of determinism is to show that determinism is not in fact one of the considerations in light of which we modify or mollify our reactive attitudes. We'll get to this very soon.

Move. (We may indeed question the assimilation with non-propositionalism and non-cognitivism, but we will not discuss this here.)

However, the non-propositionalism might be picking up on something very important; something that may not simply be stipulated away but is, quite to the contrary, to be recognised as central: the idea of a *non-rational* foundation of our conceptual scheme. This is central to Strawson's naturalism (Glock, 2022, p. 450; Heyndels, 2020, p. 102); it is closely connected to his claim that the practice *as a whole* is beyond rational justification (see chapter 2 and chapter 5). And the Reversal Move is regularly associated with the rejection of an 'external justification' (e.g., D. Shoemaker, 2022) and taken to involve a delineation of the normative framework within which the justificatory logic of our responsibility practices is contained (e.g., Beglin, 2018; Watson, 2014).

Because it is not clear quite clear what non-propositionalism is supposed to amount to, because it might pick up on something central in Strawson's approach, because it is eschewed by some and made central by others, it is hard to assess the status of (III) as a condition on the Reversal Move. Could we really have an articulation of the Reversal Move without (III)? Could we not, in the sense that it would not be a Reversal then? Or could we not, in the sense that it would not have a chance of succeeding without it?

A Possible Retraction: The Straight Path or the Short Cut?

In the last chapter, we considered an exchange between Prasad (1995) and Strawson (1995). This might have some relevance for our assessment of the Reversal Move, or at least its history.

Recall, Strawson conceded that Prasad had shown "that one of my arguments against the relevance of the thesis of determinism to the question of moral assessment and responsibility is either invalid or question-begging" (Strawson, 1995, p. 430). De Mesel has convincingly argued that Strawson takes what Prasad calls "the straight path". That is, that Strawson accepts the task of showing the irrelevance of determinism, as Prasad puts it, "directly". By Prasad's lights, this lands him in "the classical battle between the libertarians and determinists" (1995, p. 360f.). More specifically, Strawson's response to Prasad is to state what is meant by the ability to do otherwise, which he concedes is a condition on holding someone responsible (especially when one's attitude is "disapprobative" (Strawson, 1992a, p. 136))⁹³, and to

⁹³ As De Mesel (2022c, p. 5f.) has argued, Strawson is in this respect less concessive to the sceptic than Fischer's 'semicompatibilism' is (pace Sars's (2022a, n. 8) conjecture that Strawson might be said to anticipate semicompatibilism).

argue that in no relevant sense of this condition is there “an explicit or implicit denial of a thesis of determinism, stated, for example, in such simple and familiar terms as ‘Every event has a cause’” (Strawson, 1995, p. 431). He adds that,

If, in saying the above, I have dwindled into a mere compatibilist—in the company, say, of Hume—I am content with that. (Strawson, 1995, p. 431)⁹⁴

Strawson’s concession to perhaps having “dwindled into a mere compatibilist” may easily encourage the thought that he does not, in fact, rely on something like the Reversal Move to itself deliver the irrelevance of determinism. Depending, of course, on what we take the Reversal Move to be.

Strawson’s concession also suggests that, at least in ‘Freedom and Resentment’, his ambitions were greater than, or at least different from, showing the irrelevance of determinism “directly”. That he might have “dwindled”, if this really suggests that he does not endorse the more ambitious project, would still suggest that there was this more ambitious project. Furthermore, Strawson only admits that “one of my arguments” has been shown to be either invalid or question-begging.⁹⁵ Even if Strawson were himself more occupied with what seems to be (a version of) “the straight path” in his later work, to take the road less travelled by, what Prasad (1995, p. 360) calls “the detour [...] via the analysis of reactive attitudes”, might still make all the difference for the optimist project.

III. Two Reversal Theses

The Fitting Response-Dependence Thesis

We turn now to one contemporary articulation of the Reversal Move as a distinct thesis: the Fitting Response-Dependence Thesis. We’ll consider David Shoemaker’s account of this thesis as an articulation of the Reversal Move.

⁹⁴ Compare the attitude to his reply to Simon Blackburn (1998): “So we can relax: the whole issue between determinists and libertarians is an irrelevance; and the fact that it has been so long and earnestly debated is but one more illustration of the tendency of philosophers to raise a dust and the complain they cannot see. So do I emerge as a straight compatibilist? If so, *ainsi soit-il*” (Strawson, 1998e, p. 170).

⁹⁵ I thank Maria Alvarez for helping me appreciate this.

Something like it is “the most common way of conceiving of moral responsibility these days” (Coates & Tognazzini, 2013, p. 6). The common, “broadly Strawsonian” view works from an understanding of the nature of the reactive attitudes which then informs what (kind of) conditions of appropriateness for holding an agent responsible there are, taking this to in turn explain what it means to be responsible (e.g., Darwall, 2006; Macnamara, 2015a, 2015b; McKenna, 2008, 2012; D. Shoemaker, 2007, 2015; Talbert, 2008, 2012; Wallace, 1994; Watson, 2008, 2011).⁹⁶ But the Fitting Response-Dependence Thesis is not, only, that common view. As Gary Watson remarks, “very few philosophers have taken a truly Strawsonian turn” (2014, p. 16); by a ‘truly Strawsonian turn’, Watson has in mind the Reversal Move.⁹⁷ Shoemaker’s thesis is one expression of such a turn.

The Fitting Response-Dependence Thesis is typically presented in contrast to what we might be said to get on Bennett’s and Watson’s (early) readings—especially, given (III), the non-propositionalism—namely,⁹⁸

Dispositional Response-Dependence Thesis: An agent, *S*, is responsible for some action or attitude, *A*, if and only if, and in virtue of the fact that, we are disposed, under normal conditions, to hold *S* responsible for *A*.⁹⁹

Shoemaker (2017, 2022) lists three reasons for rejecting the Dispositional Response-Dependence Thesis:

- (1) *It is hard to know, even obscure, what ‘the normal conditions’ are.* To say whether a person is in fact responsible (or blameworthy) when someone is

⁹⁶ Beglin (2020, p. 2343) calls this the “attitude-based Strawsonian strategy”.

⁹⁷ In particular, “something like [Strawson’s] ‘response-dependence’ thesis” (Watson, 2014, p. 16).

⁹⁸ See Wallace (1994, p. 89), McKenna (2012, p. 34), and Shoemaker (2017, p. 496; 2022, p. 315). Rather than ‘normal conditions’, Wallace uses ‘favourable conditions’ and Shoemaker ‘standard conditions’. For now, let’s assume that this is a somewhat accurate presentation of Bennett’s reading, and not just a springboard from which others build their own accounts. McKenna (2012) argues for the Fitting Response-Dependence Thesis over the Dispositional Response-Dependence thesis but does not ultimately endorse the former as it stands, arguing, rather, that the explanation moves in *both* directions, adopting an interdependence thesis. For a minor reservation about Wallace’s adherence to the Fitting Response-Dependence Thesis, see Menges (2017, p. 17, and n. 3).

⁹⁹ Reversing the reversal, we should get the traditional approach; thus, *The Traditional Thesis (Dispositional Construal)*: We are disposed, under normal conditions, to hold an agent, *S*, responsible for some action or attitude, *A*, if and only if, and in virtue of the fact that, *S* is responsible for *A*. May this plausibly be said to represent the traditional approach?

disposed to hold them responsible (or blame them), we need to know what ‘the normal conditions’ are: What conditions are such that they “privilege some people’s actual angry responses over others in determining the blameworthy?” (D. Shoemaker, 2017, p. 497). Shoemaker finds it “entirely obscure” (2022, p. 315) what these could be and “hard to know even where to begin to answer this question” (2017, p. 497).

- (2) *The Dispositional Response-Dependence Thesis does not capture ‘the talk’*. More precisely, “the reasons to which we refer in judging someone blameworthy just do not make *justifying reference* to anyone’s dispositions; rather, they make reference to features of the blamed agent: ‘You stepped on my foot!’ [etc.]” (D. Shoemaker, 2017, p. 497, emphasis added).
- (3) *There must be the possibility of mistake (the normativity requirement)*. This is in fact two distinct, if related, worries:
 - (a) *The possibility of individual mistake*: It is an important fact that we can be *mistaken* in responding to someone as responsible and that we rationally criticise each other for holding people responsible for something they are not responsible for. In short, there is a requirement of appropriateness on holding responsible.¹⁰⁰
 - (b) *The possibility of collective mistake*: Even if a community, *G*, is sufficiently homogenously disposed to holds some agent, *S*, responsible for some action or attitude, *A*, under some conditions, this does not mean that *S* is responsible for *A*. A community may be *systematically mistaken*, such that *all* or *most* of their dispositions to hold responsible fail to fulfil the requirement of appropriateness on holding responsible.

For (*inter alia*, but in particular) these reasons, the first articulation of the Reversal Move has been rejected in favour of another thesis:

Fitting Response-Dependence Thesis: An agent, *S*, is responsible for some action or attitude, *A*, if and only if, and in virtue of the fact that, it is *fitting* to hold *S* responsible for *A*.¹⁰¹

¹⁰⁰ This objection is plausibly an inheritance from Wallace’s rejection of what he took to be Bennett’s non-cognitivist reading (see, McKenna, 2012). (See also Scanlon (1988), which informs Wallace’s discussion.)

¹⁰¹ On the reverse order of explanation, we get the *Traditional Thesis (Fittingness Construal)*: It is fitting to hold some agent, *S*, responsible for some action or attitude, *A*, if and only if,

The chief achievement of moving to the Fitting Response-Dependence Thesis is the possibility of mistake: it is not whether people *actually* (are disposed to) hold someone responsible that determines whether the person is responsible; what determines whether a person is responsible is whether it is *fitting* to hold or be disposed to hold them responsible.

This thesis of course brings with it the task of explaining the notion of ‘fittingness’.¹⁰² Beyond this, introducing ‘fittingness’ also raises issues analogous to those raised in (1) against the Dispositional Response-Dependence Thesis: What are the fit-making considerations? Why are *these* the fit-makers?

Shoemaker has recently offered an answer to this set of questions.¹⁰³ He focuses on blame, and proposes that whatever the fit-making considerations

and in virtue of the fact that, *S* is responsible for *A*. This is a much more plausible presentation of the traditional approach than the dispositional construal of that approach. What does this tell us about the history of the Reversal Move? Formulated as the Fitting Response-Dependence Thesis, the Reversal Move clearly makes contact with the Fitting Attitude Analysis of Value. Some trace this back to Kant (Suikkanen, 2009), but more often its first exponents are taken to be Henry Sidgwick (1874) and Franz Brentano (Brentano, 1889/2009); later historical advocates include C. D. Broad (1930) and A. C. Ewing (1947). (For a brief history, see Rabinowicz & Rønnow-Rasmussen (2004, pp. 394-400).) The Fitting Attitude Analysis has grown popular in recent years (see e.g., Garcia & Green Werkmäster, 2018; Howard, 2018, 2023; McHugh & Way, 2016; Naar, 2021; Suikkanen, 2009). Some recent exponents include John McDowell (1985), Allan Gibbard (1990) and T. M. Scanlon (1998). These latter three are all noteworthy in this context: McDowell was influenced by Strawson; Shoemaker (2017, p. 482, n. 2) himself notes significant tangents with Gibbard; Scanlon influenced Wallace’s (e.g., 1994, pp. viii, 74) account of responsibility. The convergence of the Fitting Attitude Analysis and self-professedly Strawsonian accounts of responsibility may furthermore be discerned, for example, in the fact that many responsibility theorists understand moral responsibility as a matter of being either blameworthy or praiseworthy, partly for the reason that “moral responsibility” is otherwise a bit suspect or mysterious (cf., Rabinowicz & Rønnow-Rasmussen, 2004, p. 400); the “worthy”-suffixed terms lend themselves much better to a Fitting Attitude Analysis than “moral responsibility” does (cf., Deonna & Teroni, 2021). Recently, this connection has been examined by Leonard Menges (2017, 2020) and, at dissertation length, by Marta Johansson Werkmäster (2023).

¹⁰² Rather than ‘fittingness’, we might wish to use ‘merit’ or ‘appropriateness’. These terms are sometimes used interchangeably. Sometimes they are used to mean *importantly* different things. We’ll use the terms interchangeably. Important question about fittingness include what fittingness includes and excludes (cf. ‘the moralistic fallacy’ (D’Arms & Jacobson, 2000; Yao, 2023) and ‘the wrong kind of reasons problem’ (Rabinowicz & Rønnow-Rasmussen, 2004)), and if fittingness is a *sui generis* notion or to be understood in terms of, for example, having (*pro tanto*, sufficient, or conclusive) reason to, in our case, hold someone responsible. We’ll not discuss these issues here.

¹⁰³ Menges (2017, p. 20) argues that the Fitting Response-Dependence Thesis at least answers the question negatively. According to the Fitting Response-Dependence Thesis, he argues,

are, these are the fit-making considerations *because* they “are the sort of properties to which *refined human anger sensibilities tend* to respond with blaming anger” (2022, p. 319, emphasis added).¹⁰⁴ As Shoemaker (2022, p. 319) himself recognises, this sounds suspiciously like the Dispositional Response-Dependence Thesis, only in terms of ‘refined human sensibilities’ rather than ‘normal conditions’.

Something, I think, is off with the dialectic of the Fitting Response-Dependence Thesis. Looking closer at this might help us get a better grasp on what the Reversal Move is supposed to be.

Let’s transpose Shoemaker’s objections against the Dispositional Response-Dependence Thesis to the more refined thesis. First, let’s ask: How can we know which are the *refined* human sensibilities, and which are not? Is it any more “obscure” what counts as ‘refined human sensibilities’ than what counts as ‘the normal conditions’?¹⁰⁵ If it not easier to know this or any less “obscure” what they are supposed to be, then we have lost one stated motivation for moving to the Fitting Response-Dependence Thesis from the Dispositional Response-Dependence Thesis.

In fact, in this respect, the situation must be *worse* for the Fitting Response-Dependence Thesis. The ‘normal condition’ clause is not *replaced* by ‘refined human sensibilities’ but remains a part of this elaboration of the Fitting Response-Dependence Thesis. The claim is that the fit-makers are the fit-makers because refined human sensibilities *tend* to react in a particular way. Even refined human sensibilities may be under distorting circumstances—i.e., not operating under normal conditions.¹⁰⁶ Thus, whether or not it is more

whatever the facts are that make it the case that it is appropriate to hold *S* responsible for *A*, the fact that *S* is responsible is *not* among them. If so, the fact that *S* is responsible for *A* is ruled out for playing any justificatory role in our practices. Compare this with the Meaning-Based Account below.

¹⁰⁴ Note that the Reversal Move and the Affective Account of blame are entangled on Shoemaker’s presentation.

¹⁰⁵ Shoemaker himself, we may note, certainly identifies some of the ‘normal conditions’ in defending his refined sensibility explanation, for it is only “when the person with the refined sensibility *has a clear-eyed view of the matter, isn’t tired or depressed, and isn’t under some other distorting influence*” (2022, p. 319, emphasis added) that her responses are apt to play the role Shoemaker’s analysis takes them to play.

¹⁰⁶ Patrick Todd (manuscript) presents Shoemaker’s development of the Fitting Response-Dependence Thesis as a new thesis, a “normative” version of the Dispositional Response-Dependence Thesis: “(NDRD) [S] is blameworthy (and thus responsible) for some action of attitude *A* if and only if, and in virtue of the fact that, *the person with the refined anger sensibilities* would be disposed to respond to [S] with blaming anger for *A* in certain standard conditions” (p. 7, original emphases). Also on Todd’s understanding, then, the view

“obscure” what counts as refined human sensibilities, if it is *the least bit* “obscure” what counts as such, then (1) is, if anything, a reason to *prefer* the Dispositional Response-Dependence Thesis *over* the Fitting Response-Dependence Thesis, rather than the other way around.¹⁰⁷

The possibility of individual mistake, (3a.), is also dubiously appealed to in motivating the Fitting Response-Dependence Thesis. That someone may be disposed to hold *S* responsible for *A* does not simply mean that *S* is responsible for *A*, according to the Dispositional Response-Dependence Thesis. Given the ‘normal condition’ clause, if someone is not under the normal conditions, their disposition to hold a person responsible does not mean that, or make it so that, the person is responsible; they are simply making a mistake in being so disposed (cf., Wallace, 1994, p. 90’, n. 11). There is the possibility of being mistaken in responsibility attributions and (pace, McKenna, 2012, p. 35; D. Shoemaker, 2017, 2022) the Dispositional Response-Dependence Thesis may account for that—at least at the individual level. (We’ll discuss (3b.), the possibility of *collective* mistake, later.)

We may also query what motivation (2) really supplies. For one thing, it is not clear that the Dispositional Response-Dependence Thesis could not provide a comprehensive (not obviously implausible) account of ‘the talk’ of the practice. Do we not often say, *simply*, that someone is responsible for something—on the assumption, as a Dispositional Response-Dependence Thesis’s talk-account might have it, that we are in normal conditions? Of course, when two or more parties are disposed to make conflicting judgements, this assumption may be upset. In such cases, what will be said is, for example, that the conflict is only apparent—that the parties’ conditions are relevantly different—or that at least one of them is not under normal conditions. Looking beyond Strawson, and “Strawsonians” about responsibility, we find such accounts.¹⁰⁸

Furthermore, it is not clear that ‘in virtue of the fact that’ should imply that the kind of explanation that either thesis supplies is a *justificatory* one; in particular, in terms of the *justificatory reasons* practitioners would or should appeal to within the practice (adumbrated, ‘the talk’). Most notably, that kind of explanation is in fact not what we get from the Fitting Response-Dependence Thesis. The relata of the thesis cannot plausibly be said to relate

Shoemaker presents is not only very close to the original Dispositional Response-Dependence Thesis but, furthermore, does include a ‘normal conditions’ clause.

¹⁰⁷ “If anything”, because it might not be a reason to prefer either.

¹⁰⁸ To take two different, but not unrelated, examples: see Wilfred Sellars (1997, §§12-18) and John McDowell (1999).

in this way on such an account: the *justificatory reason* (to which practitioners would or should appeal to in the practice) why *S* is responsible for *A* is not that *S* merits being held responsible for *A*—that is not a justificatory reason appealed to in the practice.¹⁰⁹

What the Fitting Response-Dependence Thesis supplies is rather a *metaphysical explanation*, of what *constitutes* and *determines* the conditions for responsibility (see, D. Shoemaker, 2022, p. 319). This, however, we get just as well from the Dispositional Response-Dependence Thesis, if only of a different kind (though the difference, as noted, might not be so great after all).

Finally, consider (3b.), the possibility of *collective* mistake. There is a worry here due not to the similarity between the Dispositional and the Fitting Response-Dependence Theses, but to their supposed dissimilarity. Does the Reversal Move imply that, if we were to regularly (be disposed to) hold, say, severely mentally disabled people responsible, then they would be responsible? If it does, that by itself is, according to some, a *reductio ad absurdum* (e.g., Fischer & Ravizza, 1993, p. 18; Todd, 2016). The Reversal Theorist might therefore want to resist this conclusion. However, if there is *no sense* in which this would in fact be so—if what is fitting is unvarying, holding independently of how human practices actually are—then in what sense does the concept of responsibility and what it is to be responsible *depend* on those practices? (De Mesel, 2022a, p. 1906) On a response-dependent conception of a concept, is not at least some conceptual change to be expected from such a radical change in practice as that which the challengers point to? From this perspective, the Reversal Move risks being diluted if it makes concessions to the challenge from the possibility of collective mistake.¹¹⁰

Where does Shoemaker’s elaboration of the Fitting Response-Dependence Thesis land on this? In his 2017 paper, Shoemaker says that, for those things

¹⁰⁹ Heyndels and De Mesel (2018) advance a related objection (also, De Mesel, 2022a, p. 1905). They agree (as the Dispositional Response-Dependence Thesis is taken to suggest, see (2)) that it would be strange to supply as a justifying reason for blaming someone, ‘I regard you as blameworthy because people typically do regard people like you blameworthy for things like this’. They object, however, that it would be equally strange to say, ‘I regard you as blameworthy because it is fitting to regard you as blameworthy’. Thus, they argue, if (2) is a reason to reject the Dispositional Response-Dependence Thesis, as Shoemaker argues, it is also a reason to reject the Fitting Response-Dependence Thesis.

¹¹⁰ Victoria McGeer goes so far as to say that there no “nonquestion-begging way” for the possibility of collective mistake to be levelled as a challenge against the Reversal Theorist because, if they are “thorough-going”, the Reversal Theorist’s “view clearly presupposes the very thing [the challengers’] reject” (McGeer, 2019, p. 308), viz. that what is appropriate is not grounded in and dependent on the practice. (She calls her view, not a Reversal view, but “Conventionalist View”).

that are the fit-makers, “they are so *in virtue of* their triggering our [...] sensibilities” (2017, p. 509). This is compatible with the fact that our sensibilities may change, being triggered by different things, such that this makes for a change in what the fit-makers of responsibility are. In his 2022 paper, however, the emphasis is slightly different; Shoemaker seems to exclude this possibility. He says, “we cannot, on pain of absurdity, make praiseworthiness and blameworthiness a function of the actual sentiments of praise and blame people *happen to have*. After all, children and the insane are not responsible, even if some of us respond to them with praising and blaming sentiments” (2022, p. 305). In the end, however, Shoemaker remains “neutral about what determines *refinement* in a sensibility”, whether it is determined response-dependently or -independently (2022, p. 319).

In conclusion, the reasons for elaborating the Dispositional Response-Dependence Thesis in the direction of the Fitting Response-Dependence Thesis are not clearly reasons for this move. They do not speak for preferring the latter over the former, and so the purported relative plausibility of the Fitting Response-Dependence Thesis remains unaccounted for.

We’ll discuss the prospect for a Response-Dependence Thesis to make good on Strawson’s Promise in section IV. Before that, we’ll have a look at another, rather different articulation of the Reversal Move.

The Meaning-Based Account

The latest development of the Reversal Move, proposed by Benjamin De Mesel (in individual work (2022a) and together with Sybren Heyndels (2019)), is *the Meaning-Based Account*. This draws on Strawson’s Wittgensteinian account of meaning, thus further embedding the Reversal Move in Strawson’s philosophical outlook. It is also more refined than the articulations that have come before it in that it makes use of two different explanatory relations. This allows it to grant priority in the direction of the traditional view with respect to one of these relations while asserting that priority holds in the reverse direction for the other relation. In virtue of incorporating this duality of kinds of explanatory relations, its relation to “the traditional view” is more nuanced.

The Meaning-Based Account is not easily captured in a single thesis. But the following gloss captures at least the core idea (De Mesel, 2022a, p. 1894):

The Meaning-Based Account: (i) The meaning of an expression, such as ‘*S* is responsible for *A*’ (where *S* is an agent and *A* is an action or attitude), is determined by the rules for its use. (ii) The rules for an expression’s use originate in empirical regularities which have acquired a normative status.

De Mesel presents Strawson's theory of meaning as broadly Wittgensteinian, a characterisation Strawson himself would freely grant (Strawson, 1992a, ch. 8; 2008c, ch. 4; 2008e).¹¹¹ While the account is in terms of the meaning of certain expressions, the elucidation of our concepts is the aim. Concepts are understood as abstractions from the uses of expressions: If an expression in one language is used according the same rules as another expression in another language, then the two expressions express the same concept (De Mesel, 2022a, p. 1897).

On the Meaning-Based Account, the rules for the use of an expression are not mere regularities, but they do *originate* in regularities: they “are regularities which have ‘hardened’ into norms” (De Mesel, 2022a, p. 1895; Wittgenstein, 1978, p. 365).¹¹² What it means to be responsible is not given simply by how expressions such as ‘This is your fault’ or ‘Nicely done!’ are *actually used*; it is given by how they are *to be used*—i.e., by the *rules* for their use.¹¹³ This allows for the distinction—crucial as it is (Strawson, 2008c, p. 63)—between *actual* and *correct* use. Since rules specify *correct* use, the account satisfies the normativity requirement considered above, at least at the individual level.

To elucidate the concept of responsibility, then, we need to elucidate the rules guiding our use of the relevant expressions in practice. This amounts to a form of conceptual analysis; according to Strawson, in a fundamental respect, it is the *only* way to do conceptual analysis: “the actual use of linguistic expressions remains [the philosopher’s] sole and essential point of contact with the reality he wishes to understand, conceptual reality” (2011a, p. 90).

While conceptual analysis is a matter of *discovering* the rules of our practice, the relation between the rules and the practice is not simply an epistemic relation (cf., De Mesel, 2022a, p. 1902f.; McKenna, 2012, pp. 39-41). On the Meaning-Based Account, the rules *depend* on the practice in the sense that the practice is *the context of formation* of the rules. It is not simply that we learn of the rules by observing the practice. The practice is not only the context of

¹¹¹ Snowdon and Gomes (2023) remark that, while clear in his criticism of others, Strawson “did not give a positive account of how to understand linguistic utterances”. De Mesel may be said to develop Strawson’s somewhat incomplete account in the direction of (one understanding of) Wittgenstein’s. (There are many different readings of Wittgenstein; De Mesel (2022a, p. 1895, n. 2) is explicit that his reading is inspired by Baker and Hacker (2005, 2009) and Hacker (2019).)

¹¹² De Mesel here draws explicitly on Sorin Bangu (2021). For a similar account, but from a Humean perspective, see Hartmann (2020).

¹¹³ I use “given” in order to include both ‘determined by’ and ‘understood in terms of’.

discovery. It is also the case that our rules are determined by practice. Our practices “constitute the framework within which our concepts are born and have their lives” (De Mesel, 2022a).¹¹⁴ Call the dependence relation between the practice and the concept a *constitutive* relation.¹¹⁵ The rules for the use of expressions are *constituted* by the practice, both in the sense that they have emerged from the practice—are regularities of the practice which have ‘hardened’ into norms—and that they depend on the practice for their continued existence. The practice is prior to the concept in the order of being. In this sense, the concept is not independent of the practice. Since the meaning of an expression is determined by the rules for the use of that expression, and the rules emerge from the practice, the meaning of an expression is constituted by the practice.

¹¹⁴ The practice could be otherwise, and the rules could, likewise, be otherwise. But if we ask, ‘What does “responsibility” mean?’, the answer given by stating the rules for its use is not in any way imperfect because the word ‘responsibility’ *could have* meant something else if our practices and so our rules had been different. Nor is it in any way imperfect because we might not necessarily have had *this* concept of responsibility. *Our* concept of responsibility is given by *these* rules—another concept, which may very well be another concept of *responsibility*, is necessarily, in virtue of being a different concept, given by different rules and a different practice. See discussion of the possibility of collective mistake below.

¹¹⁵ De Mesel (2022a, p. 1901) calls it *genealogical*. I refrain from adopting this term here because the term will be used for a slightly different relation in later on (in particular, in chapter 6). However, it is a bit unclear how different the two uses of the term actually are. De Mesel distinguishes between not only our *concept* and our *practice*—between which (in our terms) a constitutive relation holds, by the rules of the practice determining the concept—but also between our *practice* and our *natural concerns*. According to De Mesel, a *naturalistic* explanation “shows how our *practices* of holding responsible are rooted in *natural facts* about us and the world we live in” (2022a, p. 1899, original emphasis). A *conceptual* explanation, on the other hand, “shows how our *concept* of responsibility is rooted in our *practices* of holding responsible” (De Mesel, 2022a, p. 1899, original emphasis). Sometimes, it seems, De Mesel (2022a, p. 1901) uses the term ‘genealogical dependence’ to mean a relation that holds between natural concerns and concepts—as a kind of conceptual-cum-naturalistic explanation. Sometimes, however, De Mesel (2022a, p. 1902) seems to use the term to mean a relation that holds (only) between practices and concepts. The latter seems to be also the sense in which Bangu (2021, p. 453) uses the term. (The former, however, is the way Matthieu Queloz (e.g., 2020, 2021a, 2021b) uses the term; an inconsistency here might be due to De Mesel drawing on both Bangu and Queloz.) De Mesel is in any case clear that his “focus in [the] paper is on the conceptual strand [as distinct from the naturalistic], because my aim is to shed light on the relation between being responsible and our practices of holding responsible” (2022a, p. 1900). To avoid risk of confusion—given that we’ll use the term in the sense of conceptual-cum-naturalistic explanation later on—we’ll here use ‘constitutive relation’ instead of ‘genealogical’ in order to denote the relation between concepts and practices (the kind of conceptual explanation De Mesel outlines).

The contrast with the traditional approach, then, is this: on the traditional approach, the concept is prior to the practice and, therefore, so are also the rules for the use of the relevant expressions; therefore, to learn the rules for the use of the relevant expressions (e.g., those involving the concept of responsibility), we do not look to the practice, but we look to the concept, to its inherent meaning.

While the difference between the traditional and the Reversal approach is one of method—how to elucidate the concept, how to determine the rules of the practice—this difference in method is grounded in a difference in the order of being: whether the concept or the practice is taken to be prior.

The Meaning-Based Account is itself silent on what the rules of the practice are. That the rules—and so the meanings of our expressions and our concepts—are constituted by the practice does not imply that a rule in our practice states, for example, ‘It is appropriate to hold someone responsible for an attitude or action because we have treated and still do treat some fact as relevant for whether people are responsible, to the effect that this fact now has the status of a rule’ (De Mesel, 2022a, p. 1905).¹¹⁶ The Meaning-Based Account does not say *what* the rules of our practice for the use of certain expressions are, only that it is the rules (whatever they are) that determine the meaning of our expressions and, likewise, what our concepts are.

This means that it is open for the Meaning-Based Account to hold, plausibly enough, that it is a rule of our practice that *it is appropriate to hold someone responsible only if the person is responsible*.¹¹⁷ If this is correct, to hold someone responsible is *justified* only if they are responsible. In this sense, (the appropriateness of) holding responsible *depends* on the agent being responsible. Call this dependence relation between holding responsible and being responsible a *justificatory* relation.¹¹⁸

Critics of the Reversal Move advance this deeply intuitive conceptual connection—that *holding* responsible is justified only if the agent *is* responsible—as an objection to the Reversal Move (Brink & Nelkin, 2013;

¹¹⁶ The Meaning-Based Account, almost by definition, captures ‘the talk’ of the practice (cf. (2) above).

¹¹⁷ It seems like Strawson would agree: “It is quite true, indeed tautological, that an agent can be justly blamed for an action only when he can be held morally responsible for it; and it is true that he is responsible for so acting only if he acted freely” (Strawson, 1998e, p. 170). There is a gap between ‘can be held morally responsible’ and ‘is responsible’ that needs to be bridged, but the swiftness of Strawson’s movement here suggest that this gap requires no leap, but is straightforwardly bridged.

¹¹⁸ If there is an order of priority here, it is in *the space of reasons* that being responsible may be said to be prior to holding responsible.

Fischer & Ravizza, 1998). But it is not an objection to the Meaning-Based Account, since this thesis does not say anything about *what* the meaning of ‘responsibility’ is.¹¹⁹

Distinguishing between *justificatory* and *constitutive* dependence, the Meaning-Based Account may account for the fact that the concept of responsibility is *constitutively dependent* on the practice of responsibility, while accommodating the thought that holding someone responsible is *justificatorily dependent* on the person being responsible.

Finally, how does the Meaning-Based Account respond to (3b.), the challenge from the possibility of collective mistake? De Mesel & Heyndels (2019) argue that the Meaning-Based Account does not imply that if a group, *G*, would systematically hold some (set of) agent(s), *S* (*/S/*), responsible for some action or attitude, *A*, then *S* (*/S/*) would be responsible.¹²⁰ More specifically, *S* (*/S/*) might not be responsible *in our* sense, according to *our* concept of responsibility. Rather, since the rules of the practice of *G* would be radically different from our rules, *G* would not have our concept of responsibility, but a different concept. Even if the members of *G* use the word “responsible” in describing the relevant agent(s), which we might not, this does not mean that we disagree on whether they *are* responsible—if our use of “responsible” in this very sentence is taken to express *our concept* of responsibility.¹²¹ On the Meaning-Based Account, then, we can concede that there is something absurd about saying that, for example, some severely mentally disabled people are responsible, because, given that the rules for the use of this expression are as we take them to be, this would be a flagrant misapplication of that concept.

¹¹⁹ Menges (2017, p. 18f.) argues that it is not an objection to the Fitting Response-Dependence Thesis that being responsible is a necessary condition on its being appropriate to hold the person responsible, because if the appropriateness of holding the person responsible makes it the case that she is responsible, then it is impossible for it to be appropriate to hold the person responsible for something which she is not responsible for. However, being a necessary condition falls short not only of being a justification, but also of being a necessary condition *on justification*. Menges is right that, according to the Fitting Response-Dependence Thesis, whenever it is appropriate to hold someone responsible, it is necessarily the case that the person is responsible. But this still does not make sense of the *conceptual* connection that is at issue, between the appropriateness of holding responsible and being responsible.

¹²⁰ See also De Mesel (2022a, p. 1908) and, for a similar response, Bengtson (2019).

¹²¹ Whether *G*’s concept should be considered a *conception of responsibility*, such that both it and our concept are responsibility-concepts, will plausibly depend on further particularities of the case.

IV. Strawson's Promise and the Reversal Move

Without adjudicating which of the two considered articulations of the Reversal Move is the most plausible, interpretatively or independently, their respective prospects for making good on Strawson's Promise may still be assessed.¹²²

In his criticism of the Reversal Move, Todd argues that, if the Reversal Theorist is to get an anti-pessimistic conclusion, they have to deny "that the fairness [or, more generally, the appropriateness] of holding an agent responsible is grounded in facts—and here it would seem to be *any* facts—about the agent at all" (2016, p. 223, original emphasis). But neither of the Reversal Theses considered denies this—nor should they. The grounding relation at issue here is one of justificatory dependence (that, if holding an agent responsible is to be justified or appropriate, certain facts about the agent must hold). Even if, as on the Fitting Response-Dependence Thesis, the fact that the agent is responsible is grounded in the fact that it is appropriate to hold the person responsible, the fact that it is appropriate to hold the person responsible is itself grounded in other facts, which plausibly (in the case of responsibility) are facts about the agent (Menges, 2017).¹²³

Todd's thought is that if we allow for facts about the agent to be relevant, then the pessimist can argue that the relevant facts (e.g., that the agent could have acted otherwise) are such that they require the falsity of determinism. What needs to be shown is that the facts that ground the appropriateness of holding responsible (the facts partly constituting or implied by whatever the fit-makers or the rules are) are not such that they require the falsity of determinism. If it is to be anti-pessimistic, a Reversal Thesis must somehow exclude the possibility that some condition on responsibility would require the falsity of determinism.¹²⁴ How can this be done?

¹²² Even if we grant that the Meaning-Based Account is plausible as an interpretation of Strawson's view, this need not imply that the Fitting Response-Dependence Thesis is not—it might be a mistake to view them as relevantly competing views. It is however relevant to note that, at least by Shoemaker's lights, the *Fitting* Response-Dependence Thesis does not purport to be an interpretation of Strawson, but to move beyond Strawson's view (which is perhaps, on this story, to be understood as a *Dispositional* Response-Dependence Thesis).

¹²³ Without a further explanation of what the relevant facts are, an account of this kind (closely akin to a Fitting Attitude Analysis of Value) may already *demystify* the notion of being responsible by explaining that notion in more familiar terms, i.e., in terms of certain attitudes being appropriate (see, Rabinowicz & Rønnow-Rasmussen, 2004, p. 400).

¹²⁴ Following Beglin's terminology, there are two senses in which a Reversal Move may *exclude* this possibility: a formal and a substantive sense. A Reversal Thesis that does rely on facts for the justification of holding people responsible, in the sense we've seen that both the Fitting Response-Dependence Thesis and the Meaning-Based Account does, may not be

Of course, it can be shown, as Prasad puts it, “directly”. That is, we can take “the straight path” and provide a conceptual analysis of the concept of responsibility that shows that, among the conditions for appropriately applying this concept in any particular case, there is, as Strawson puts it, neither “an explicit [n]or implicit denial of a thesis of determinism” (1995, p. 361). As noted, in his later writings, especially ‘Freedom and Necessity’ (1992a, ch. 10), Strawson is explicit that this is how he proceeds.

Prasad notes another way of showing the irrelevance of determinism—viz., “the detour, [Strawson] makes via the analysis of reactive attitudes” (1995, p. 360). However, by Prasad’s light, “it would take away much of the virtue of” that detour if the irrelevance of determinism (still) has to be shown “directly” (ibid.).¹²⁵ Of course, we need not agree in that assessment.

If we are to not take the “straight path”, sooner or later—if “the detour” is to be “a viable short cut” (Prasad, 1995, p. 360)—then what does this analysis of the reactive attitudes have to be like, and what does it have to show? Can the Reversal Move make good on Strawson’s Promise without, eventually, relying on the accuracy of an elucidation of the relevant conditions for responsibility (whether in terms of fit-makers or rules for the use of certain expressions)? If not, has not the revolutionary air waned? Has it not then failed to circumvent the traditional debate? We are returned to the question: What is the Reversal Move supposed to be?

In this light, the Meaning-Based Approach might seem inadequate. The task of elucidating the rules for the use of certain expressions—reformulated, with respect to responsibility: articulating the conditions for when it is appropriate to hold someone responsible—is just what it has us doing. There is no guarantee that our analysis will not show that our concept of responsibility lacks application if determinism is true. We must take “the straight path”. Does accepting the Meaning-Based Account therefore leave everything as it was? No. That our practices “constitute the framework within which our concepts are born and have their lives” should shape our analysis.¹²⁶ It encourages

“*formally* antilibertarian”. It is in principle possible that the relevant facts are such that determinism is not irrelevant. However, a Reversal Thesis may still be “*substantively* antilibertarian”. A Reversal Thesis will be substantively antilibertarian, in Beglin’s sense, if it “give[s] us *prima facie* reason to doubt libertarian views of moral responsibility” (2018, p. 615). I take this to be a minimum requirement. A more ambitious Reversal Thesis holds that, while it is not formally impossible, as a matter of fact, a libertarian view of responsibility is mistaken; this is still a substantively antilibertarian argument, I take it, only more ambitious.

¹²⁵ Todd (manuscript) effectively echoes this sentiment.

¹²⁶ For example, Pereboom distinguishes two major routes to compatibilism; one is Strawson’s, that determinism is irrelevant because our reactive attitudes are what constitute responsibility

certain question, rather than others.¹²⁷ That is as real a philosophical contribution as there is. However, it remains true that, for the same reason that the Meaning-Based Account may easily accommodate the justificatory dependence of holding responsible on being responsible, because it is itself silent on what exactly the rules are, it does not by itself deliver on Strawson's Promise. It might therefore, as a Reversal Move, seem inadequate.

The prospects for a Response-Dependence Thesis to make good on Strawson's Promise depend on there being a naturalistic explanation of how the concept of responsibility is determined by human sentiments such that the conditions for that concept's application are shown not to include any condition dependent on the falsity of determinism. In response to Shoemaker's elaboration, it is open for the pessimist to claim (perhaps aided by a Spinozistic story of our sense of freedom) that refined human sensibilities, if truly refined, would not be disposed (when not under distorting conditions) to hold someone responsible if (it is believed that) determinism is true. If this is correct, then, by Shoemaker's own account, it will not be fitting to hold someone responsible if determinism is true. So far, the Response-Dependence theorist has not excluded this possibility.¹²⁸ Without such an account, we have at best only postponed taking "the straight path". A Response-Dependence Thesis may prove "a viable short cut" only if its naturalistic explanation is such that it itself rules the relevance of determinism out.

and provides the basis of that concept; the other "and more common aims to differentiate causal circumstances of actions that exclude moral responsibility from those that do not"—a traditional compatibilist project—and this "route to compatibilism is typically developed by surveying our intuitions about blameworthiness and praiseworthiness in specific kinds of examples" where these intuitions "are then employed to motivate conditions [...] required for moral responsibility" (2014, p. 71). On the Meaning-Based Account, even if our compatibilist project is essentially the traditional one, it would not take the form that is "more common".

¹²⁷ A paradigmatic example of this, which De Mesel (2022a, p. 1897f., also, pp. 1896, 1904 n.8) incorporates in his account, is the Wittgensteinian idea that "Essence is expressed in grammar" (Wittgenstein, 2009a, §371), why we should focus on the grammar—i.e., the rules for the use of the relevant expressions—rather than asking 'What is X' which is also to ask "for the clarification of a word; but [...] makes us expect a wrong kind of answer" (Wittgenstein, 2009a, §370). For further discussion, see De Mesel & Heyndels (2019, p. 803).

¹²⁸ Relevantly, Shoemaker (2022, p. 319) himself concedes that, for all that he has said, it might be that 'refined' should be understood in response-independent terms.

V. Conclusion

What is needed if a Reversal Thesis is to deliver on Strawson's Promise, I think, is a naturalistic explanation that renders it plausible that determinism would not be relevant for responsibility. The Fitting Response-Dependence Thesis must embrace their naturalistic strain, and the Meaning-Based Account must be supplemented with a (deeper) naturalistic explanation at the level between human nature and actual practice.¹²⁹ What is needed is an account of the natural underpinnings of our conceptual scheme which gives us reason to think that a condition incompatible with determinism would not be one of our conditions for responsibility.¹³⁰ If any Reversal Move is to make good on Strawson's Promise, it must include such a naturalistic explanation. In chapter 6, we will consider a candidate for such an explanation—a pragmatic genealogy—and assess its fit with respect to Strawson's approach. I'll argue that, suitably developed, we can deliver on at least a version of Strawson's Promise if we adopt such a naturalistic account.

There is another way of dealing with the sceptic, however, which Bennett, Watson and Prasad all remark on: the Inescapability Claim. None of them is very impressed by it. That impression has been lasting. Nevertheless, in light of what has come before, we may now be ready to reconsider this route. In the next chapter, I argue that, when properly understood, the Inescapability Claim holds the key to making good on Strawson's Promise.

¹²⁹ De Mesel contrasts the Meaning-Based Account with naturalistic accounts, and while "naturalistic views [...] focus on the relation between concerns and practices, [...] they do not have much to say about the relation between practices and concepts" (2022a, p. 1909). The point I'm making with respect to the Meaning-Based Account is that if we are to show that our concept of responsibility does not require the falsity of determinism, then our explanation must reach beyond our practices, to our nature.

¹³⁰ Examples of accounts that purport to do precisely this are Gary Watson's (2014) Normative Framework Argument and David Beglin's (2018, 2020) Concern-Based Account. Some reservations regarding Watson's account are raised in the next chapter. In chapter 7, we will consider some points of contrast between Beglin's account and the naturalistic explanation developed there.

5. The Inescapability Claim

I. Introduction

We are now to explore the possibility that the least celebrated aspect of Strawson's response to the responsibility sceptic might in fact be the most promising way of moving beyond the traditional debate. The thesis to be developed is that the Inescapability Claim provides the route through which we may make good on, at least a version of, the promise of Strawson's approach—of affirming the fundamental irrelevance of determinism and moving beyond the traditional debate. None of the alternative routes considered—of a transcendental argument (chapter 2), of the Statistical Reading (chapter 3), or of the Reversal Move (chapter 4)—have shown themselves capable of making good on Strawson's Promise.

At this point, however, doubts may surface. Has that elusive sense of Strawson's Promise been, from its very inception, illusory? The claim defended here is that it has at least not been wholly illusory. In a way, quite to the contrary: the claim is that the nature of the promise has as of yet not even been fully appreciated.

In this chapter, we'll start (in section II) by characterising what the Standard Reading of the Inescapability Claim is in light of how it has been presented in the secondary literature. The object of this chapter is *not* to argue directly against the claims of the Standard Reading. The object is rather to replace that reading with an interpretatively more adequate understanding of Strawson's claim. We'll begin by refocusing the Inescapability Claim (in section III), treating the claim as one among other affirmations of framework commitments that the descriptive metaphysician seeks to lay bare. Proceeding, we will seek to explicate *what* precisely is said to be inescapable (in section IV) and in what *sense*, or *senses*, it is inescapable (in section V).

The central argument of this chapter is that Strawson's view is a *Minimal Optimism*. On this reading, what is said to be inescapable is the *concept* of responsibility, not any particular *conception* of responsibility. The concept is inescapable in two different senses: it is *practically necessary* for the existence of human society, given the needs and situations of human beings; it is

conceptually necessary if we are to have anything that we can make intelligible to ourselves as a system of human relationships.

At the end of this chapter (section VI), the practical necessity will be seen to be in a sense more basic than the conceptual necessity. This mimics Strawson's own presentation. While this reading pivots on an assimilation of the commitment to responsibility with that to other framework features of our conceptual scheme—which allows us to cast some light on what the status of the claim is and, more so, on what kind of claim it is (in section V)—Strawson recognises a “twist or complication” in the case of responsibility, a disanalogy to the rest. This requires us to further qualify the Inescapability Claim. It is at this point that Strawson relies on the practical necessity of involvement in interpersonal relationships and on the practical necessity of the concept of responsibility.

The reading presented here differs significantly from the Standard Reading of the Inescapability Claim. This is a good thing: the Inescapability Claim as standardly understood is not well-received. By replacing the Standard Reading, this interpretation is rehabilitative of Strawson's approach. The rehabilitation consists partly in rendering inapplicable the standard criticisms and partly, simply, in offering on Strawson's behalf a more nuanced and plausible response to responsibility scepticism.

II. The Standard Reading

The Standard Reading of the Inescapability Claim

Let's start with a paradigmatic expression of the Inescapability Claim from 'Freedom and Resentment':

The human commitment to participation in ordinary inter-personal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as inter-personal relationships as we normally understand them; and being involved in inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question. (Strawson, 2008a, p. 12)

This is the reason why Strawson is “strongly inclined to think that it is, for us as we are, practically inconceivable” (2008a, p. 12) that we find ourselves in

the sceptical scenario of universal objectivity of attitude—i.e., in the complete absence of any practice, concept, or attitudes of responsibility.

In the secondary literature, a particular understanding of the Inescapability Claim is widely assumed. Consider, for example, the following presentations of Strawson's Inescapability Claim:

According to Thomas Nagel, the claim is that “it is not possible for the reactive attitudes to be philosophically undermined *in general* by any belief about the universe or human action, including the belief in determinism” (1986, p. 124, original emphasis).

According to Paul Russell, the claim is that “it is psychologically impossible to suspend or abandon our reactive attitudes entirely” (2017d, p. 33).

According to Derk Pereboom, the claim is “the psychological thesis that our reactive attitudes cannot be affected by a general belief in determinism, or by any such abstract metaphysical view, and that therefore the project of altering or eliminating our reactive attitudes by a determinist conviction would be ineffectual” (1995, p. 37; also, 2001, p. 92; 2014, p. 154).

According to András Szigeti, the claim affirms “a thoroughgoing psychological incapacity rooted in human nature which makes it impossible for us to give up the practice of responsibility-attributions” (2012, p. 103).¹³¹

According to Gary Watson, “[t]he psychological incapacity claim is that we (as we are now) could not be led to abandon that framework, *whether or not it is correct*” (2014, p. 25, original emphasis).

The Standard Reading involves treating the claim as a *psychological* thesis, according to which our responsibility practices are *immune* to the potentially disruptive force of believing that determinism is true.¹³² So understood, the Inescapability Claim implies that our reactive attitudes “cannot be affected”,

¹³¹ Szigeti distinguishes between (at least) four different arguments from inescapability; the Standard Reading is just one reading that he offers.

¹³² See also, for example, Fischer (2014, p. 104f.), McKenna (2005, p. 164), and McKenna & Russell (2008a, p. 6). Sars (2022a) distinguishes between two claims or arguments: ‘Incapacity’ and ‘Inconceivability’. The former is the Inescapability Claim as Standardly understood—that, due to our psychological capacities, “we are just stuck with our ordinary interpersonal relationships” (p. 83). The latter is interestingly, or perhaps misleadingly, described as presenting a “practical impossibility for beings like us” (p. 83), consisting at least in part of a “global cognitive dissonance” (p. 85), but is also a “conceptual argument” (p. 83).

not by (theoretical, sceptical, abstract) arguments at least—a conviction of the truth of determinism would be “ineffectual”. So we have the reactive attitudes “whether or not” our practices are “correct”. The reason for this cognitive or otherwise psychological immunity is taken to be that the reactive attitudes are “rooted in human nature” such that, “for us”, it is “psychologically impossible” to give them up.

In understanding the Inescapability Claim as a *psychological thesis* (e.g., Miller, 2014a, 2014b; Pereboom, 1995; Russell, 2017d; Sziget, 2012; Watson, 2014), the Standard Reading assumes, for one thing, that to abandon the practice, concept, and attitudes of responsibility is “psychologically impossible” (Russell, 2017d, p. 33)—i.e., that the necessity is a kind of psychological necessity—and, further, that this is impossible because of a “psychological incapacity” (Sziget, 2012, p. 103)—i.e., that the necessity holds in virtue of “our psychological make-up” (Coates, 2017, p. 804).¹³³ A third component of the Standard Reading may be noted, which coheres nicely with the psychological construal: an individualistic assumption. That we are inescapably committed to the practice of responsibility is taken to mean that *each and every one of us* are so committed. It is in virtue of individual human psychology that the practice, concept, and attitudes are inescapable and, correspondingly, the necessity here involved is a kind of *psychological necessity*, an inescapability of individual human beings.¹³⁴

The Standard Criticisms of the Inescapability Claim

On the received view, the Inescapability Claim plays a central role in Strawson’s argument. Because of the centrality of this claim, Strawson’s

¹³³ Heyndels (2020) correctly argues that “the Psychological Impossibility Argument”, as he calls it, is not Strawson’s argument because such a *justificatory* project is not Strawson’s project (cf., Campbell, 2017; De Mesel, 2018). However, he still claims that a central premise of that argument—which is the Inescapability Claim—is Strawson’s claim: “Strawson indeed believed that it is, *in some sense*, impossible to abandon our responsibility practices” (2020, p. 105). This is right. But in what sense he thought it was is not specified on Heyndels’s account. Given that it is *as a premise* of the *Psychological Impossibility Argument* that Strawson is said to affirm it, the reader is left to infer that the claim attributed to Strawson is that it is *psychologically* impossible to abandon our responsibility practices. I doubt that this is actually Heyndels’s view. The point here is only that previous rejections of the Psychological Impossibility Argument as Strawson’s argument have still not said enough; more needs to be said in order to replace the Standard Reading.

¹³⁴ It is natural to assume that the kind of reflection that is supposed to deliver this insight is the study of psychology.

approach is often met with some dissatisfaction, even by those otherwise sympathetic. The Inescapability Claim is generally taken to be, in some sense, either,

- (i) implausible (Coates, 2017; Pereboom, 2001; Russell, 2017d; Sommers, 2007),¹³⁵
- (ii) inadequate (Russell, 2017d; Smilansky, 2008), or
- (iii) irrelevant (McGeer, 2014; Sars, 2022a; Watson, 2014).¹³⁶

This taxonomy of criticisms allows for further distinctions between versions of each of them, and this will be highlighted when relevant.¹³⁷ Differences aside, as Russell (2017c, p. 98) observes:

There is, nevertheless, a general consensus among both followers and critics alike that there are significant strands in Strawson’s specific naturalistic arguments that are implausible and unconvincing and that some “retreat” from the original strong naturalistic program that he advances is required.

This consensus, however, is itself due to that other consensus: the Standard Reading of the Inescapability Claim. All of these criticisms—but for one, the inadequacy criticism (to be considered in chapter 7)—are premised on the Standard Reading.

¹³⁵ Wallace (1994, p. 31f.) too rejects the claim as implausible but does so explicitly from the assumption that we adopt a “narrower” construal of the reactive attitudes; for that reason, his criticism is knowingly not quite of Strawson’s claim, but of a kind of development of Strawson’s actual claim.

¹³⁶ Callanan (2011), Putnam (1998), Heyndels (2019), and Glock (2022) also take it to be *irrelevant* (at least) for the anti-sceptical project (however, see Heyndels (2020)). For Callanan, this is an interpretative claim, and so it seems to be on Heyndels’s (2019) account; Glock (2022) is clear that he affirms this independently of interpretation, taking the interpretative issue to be equivocal. Putnam affirms its irrelevance not as an interpretative matter but assigns to Strawson an incoherence or tension with respect to this issue. See chapter 2 for discussion.

¹³⁷ Among criticisms in the literature, we may identify two subcategories of each kind of criticism: for (i), the implausibility criticism, there is (a) that the claim is empirically implausible, and (b) that the claim is naturalistically implausible; for (ii), the inadequacy criticism, there is (a) that the claim is inadequate for defending an ethical outlook, and (b) that the claim is inadequate for addressing scepticism; for (iii), the irrelevance criticism, there is (a) that the claim is dialectically irrelevant, and (b) that the claim is rationally irrelevant.

Take the perhaps most common criticism, a version of (i): *the empirical plausibility challenge*.¹³⁸ This objection casts doubt on the strength of the Inescapability Claim, demanding that it be significantly weakened. It insists that, while it might indeed be very hard to rid ourselves from the sense of freedom and our disposition to have reactive attitudes, this is only *difficult*—we have no reason to think that is in fact *impossible*. It might take time and effort, sure. But this does not make it an absolutely *irremovable* feature of human life.¹³⁹

Tamler Sommers (2007), for example, invites us to imagine Sally. Sally is convinced of the truth of free will scepticism and the irrationality of the reactive attitudes. Sommers then asks us: “To what degree is it psychologically possible for *her* to abandon a belief in [robust moral responsibility] and view everyone exclusively in an objective way? [...] Will Sally ever be able to feel *in her gut* that robust moral responsibility is a fiction? Will she ever really be able to take the objective attitudes towards everyone at all times?” (2007, p. 336, original emphasis).¹⁴⁰ It might be very difficult for Sally to get to the point at which she does “feel *in her gut* that robust moral responsibility is a fiction”. Indeed, people are likely to experience recalcitrant emotions for some time if they try to abandon the framework, as is a familiar fact about the dispositional roots of any evaluative practice. Sommers’s objection is only that it is too strong to claim that it is *impossible* for us to shed the reactive attitudes. If Sally *can*, if only with significant time and effort, rid herself of the reactive attitudes, the Inescapability Claim is false.

In fact, it need not even be that *Sally* can succeed in shedding her reactive attitudes if the claim is to be false. While it might not be possible for her, given her social circumstances and upbringing, it might still be the case that, if the social circumstances and conditioning of future individuals are different enough, then, eventually, *someone* may come to maturity without any reactive attitudes, without a belief in (robust moral) responsibility. If we admit that this is possible, we have to admit that the Inescapability Claim is too strong—as a claim about human psychology, it is empirically implausible.

¹³⁸ This label is adopted from Sars (2022a).

¹³⁹ This may be cast either as the possibility of adopting an exclusively objective attitude (e.g., Sommers, 2007) or of adopting attitudes that are, while not “objective”, still not “reactive”—i.e., attitudes that are not responsibility-entailing (e.g., Pereboom, 2014, 2021). In pointing to figures like King and Gandhi, Watson (2008) is getting at the same point as Pereboom.

¹⁴⁰ Sommers uses the term ‘robust moral responsibility’ to refer to the kind of responsibility that free will sceptics are sceptical about, leaving it open that there might be kinds of responsibility that are not incoherent or incompatible with a naturalistic outlook.

At this point, what is interesting for us to note about this objection is *the test* it puts forward: if some individual can live without reactive attitudes or a concept of responsibility, then the Inescapability Claim is false. That this is the test shows *the individualistic assumption* of the Standard Reading. What the Inescapability Claim, so understood, affirms is a psychological impossibility *for each and every one of us*. This reading owes to a quite general ambiguity with respect to the ‘we’ under consideration—i.e., of the claim that *we* inescapably have a practice of responsibility.¹⁴¹ There is more than one respect in which this ‘we’ is ambiguous, and we’ll get back to this more than once. We’ll return to the particular ambiguity that underlies the individualistic assumption at the end of this chapter.

Against the Consensus

In so far as the Standard Reading indeed encourages the sense that Strawson takes us to be *simply stuck with* our reactive attitudes and our practice of responsibility in virtue of our human nature, the dissatisfaction is understandable. If this entrapment in a practice is furthermore supposed to be *due to a psychological incapacity of individual agents*, the dissatisfaction is even more understandable. And, even if we were simply stuck with our practice of responsibility because of our psychological make-up, this surely would make for a feeble attempt to refute sceptical doubt. But none of this is correct.

Never is this understanding of the claim argued for. It is just widely assumed. And it is widely assumed despite the fact that nowhere, in ‘Freedom and Resentment’ or elsewhere, does Strawson put the Inescapability Claim in terms of individual human psychology. That Strawson appeals to ‘human nature’ does not itself imply that the Inescapability Claim is “a psychological thesis”; an account of human nature merely in psychological terms is simply inadequate.

If it were to be argued for, proponents of the Standard Reading might cite Strawson’s recurrent claim, for which he draws on Hume, that, just as we have a “natural disposition to belief” in material objects, we have a “natural disposition to” reactive attitudes (e.g., Strawson, 2008c, p. 31). They might also want to add that, *in precisely that passage cited above, which is taken as a paradigmatic expression of the Inescapability Claim*, Strawson explicitly

¹⁴¹ Strawson often puts the Inescapability Claim in terms of what *we* cannot do—e.g., “Being human, *we* cannot, in the normal case, [... take the objective attitude] for long, or altogether” (2008a, p. 10, emphasis added).

rules out the sceptical scenario because “[t]he human commitment to participation in ordinary interpersonal relationships is [...] *too deeply rooted and thoroughgoing*”. But this is to take the commitment to be basic in the sense that it is psychologically well-entrenched and incorrigible; to be a commitment too deeply rooted in us for any abstract arguments to be able to upset it. This may award a sense of (relative and or subjective) certainty befitting W. V. Quine’s (1969, ch. 3) project of a naturalized epistemology (cf., Glock, 2016, p. 276), but that is certainly not Strawson’s way with scepticism (Strawson, 2008c, pp. 8, 15-16, 72).

In his “Reply to Ernest Sosa” (1998c; cf., Sosa, 1998), Strawson clearly rejects the understanding of his claim assumed by the Standard Reading:

It is not merely a matter of dismissing the demand for a justification of one’s belief in a proposition on the grounds that one can’t help believing it. That would be weak indeed. (Strawson, 1998c, p. 370)

In the rest of this chapter and the following chapter, we’ll seek to surpass the Standard Reading. According to Strawson, we are not *simply stuck with* our reactive attitudes or our practice of responsibility. The inescapability of the reactive attitudes, and of the practice and the concept of responsibility, is not a matter of *psychological necessity*. And to affirm the inescapability of the practice, concept, and attitudes of responsibility is not meant as a *refutation* of scepticism—even if it is meant to take us, in a sense to be specified, beyond scepticism about responsibility, to a kind of optimism.

III. Refocusing the Inescapability Claim

In an interview in 1992, summarising his view on the issue of free will and responsibility, Strawson puts his basic thought as follows:

[...] just as we are naturally committed, and inescapably committed, to belief in the material world and other people, so we are naturally committed to certain kinds of reaction to other people’s behaviour and our own, which imply our readiness to take disapprobative or approbative attitudes both to other people and our own. And these wouldn’t make sense unless we credited them with knowing what they were doing and responsibility for their actions. And insofar as freedom is linked to responsibility, then indeed we are naturally committed to the belief in freedom of action. [...] this doesn’t involve some metaphysical, and indeed unintelligible, libertarian conception of freedom ... but it does

involve all that we actually need, or can hope for, in the way of responsibility and freedom. (Strawson et al., 2008, p. 93f.)

Two things about this passage are particularly important to note, helping us to refocus the Inescapability Claim.

First, note the last bit, after the pause. It is from this that we'll take our cue, as it captures the optimistic attitude underlying Strawson's Promise—that we have “all that we actually need, or can hope for, in the way of responsibility and freedom”. Borrowing a phrase from Bernard Williams (1995a, p. 19), let's cast this optimistic claim as saying that *the will is as free as it needs to be*.¹⁴² This is only a *minimally* optimistic claim. It might be that the will is a lot less free than we thought it had to be, and still be all that we actually need—or can hope for. Let's call this view *Minimal Optimism*. The fundamental source of this optimism is a recognition of the necessity of responsibility. The crux of Minimal Optimism, however, and particularly of reading Strawson as proposing this view, is that of properly delimiting the Inescapability Claim, capturing the minimality of the optimism it supports, without thereby losing sight of the undoubtedly great ambition of Strawson's approach, of shoring up the fundamental irrelevance of determinism.

The second thing to note in the passage above is that the claim that the will is as free as it needs to be is presented as a claim of the same sort as the claim that there is a material world and that there are other minds; or, to put it more carefully, that we necessarily have a conception of the material world and of other minds. As argued in chapter 2, Strawson's way with the responsibility sceptic is fundamentally the same as his way with other forms of scepticism. Strawson understands our natural commitment to the material world and to other minds as being of the same kind as our natural commitment to the practice, concept, and attitudes of responsibility. If we are to understand Strawson's approach, the combination of naturalism with the project of descriptive metaphysics must be recognised. We will exploit this connection in several respects in this chapter.

One important upshot of this connection is that it helps us see that the idea that *the will is as free as it needs to be* should be understood as expressive of a *framework commitment*. In section V, we will see that there are two sides to framework commitments, and also that they are expressive of two corresponding kinds of necessities. This adds complexity to the account of the

¹⁴² The point of this phrase cannot be said to be exactly the same for Strawson as for Williams. Nevertheless, the commonality is, so I hope to show, illuminating. For a discussion of Strawson and Williams, see Russell (2024) and, on some methodological similarities, see Jubb (2016). Further similarities will emerge, particularly in the following two chapters.

anti-sceptical potential of Strawson's naturalism presented in chapter 2. This complexity is much needed because, as we go on to consider in section VI, the case of responsibility is indeed more complicated than the assimilation with the other framework commitments, to material object or other minds, even if significant and illuminating, may otherwise lead us to think.

Another important insight that we gain from seeing the connection to descriptive metaphysics concerns *what* precisely is said to be inescapable—the central claim of Minimal Optimism; what makes it *minimal*.

IV. Delimiting the Inescapability Claim

A key passage for understanding *what* is inescapable according to Strawson is when, in the penultimate paragraph of 'Freedom and Resentment', he considers our growing "historical and anthropological awareness" about the reactive attitudes' "development and [...] the variety of their manifestations" (2008a, p. 26). In consideration of such awareness, he says:

This makes one rightly chary of claiming as essential features of the concept of morality in general, forms of these attitudes which may have a local and temporary prominence. No doubt to some extent my own descriptions of human attitudes have reflected local and temporary features of our own culture. But an awareness of variety of forms should not prevent us from acknowledging also that in the absence of *any* forms of these attitudes it is doubtful whether we should have anything that *we* could find intelligible as a system of human relationships, as human society. (2008a, p. 26, original emphasis)

In consideration of conceptual variation, historically and cross-culturally, Strawson clarifies the *limits* of the Inescapability Claim: what we cannot do without is *some* form of these attitudes.

We do recognise that the form our particular network of reactive attitudes, our particular conception of responsibility, and our particular responsibility practices take is not the only form that they have taken through history or across cultures, and so not the only form they *can* take. What is inescapable is not the practice, concept, and attitudes of responsibility *in the form that we have* them, where 'we' is understood in a *contrastive sense*: a 'we' in contrast to some other 'we'. Neither is the form in which others, a 'they', or some past 'we', have it in, or have had it in, inescapable.

What is said to be inescapable is rather that we—in an inclusive sense—have *some* form of the practice, concept, and attitudes of responsibility. If we

have *any* practice, concept, and attitudes of responsibility, that is genuinely *of responsibility*—i.e., a *variation on* or *form of* such a practice, concept, and set of attitudes—then we have *some*, and that is all that is inescapable.¹⁴³

Importantly, the passage above suggests that in so far as Strawson’s “descriptions of human attitudes have reflected local and temporary features of our own culture”, the analysis is falling short of its ideal, is in some measure missing its mark (cf., Russell, 2024, p. 247f.). It is all too easy to take our own outlook for granted, failing to see the ways in which it is peculiar to us; Strawson here corrects for this, however abstractly: to the extent that Strawson’s description has reflected our (his?) socio-historically particular conception of these attitudes, to that extent is his description of these attitudes, by his own standards, imperfect.¹⁴⁴

The clarificatory passage at the end of ‘Freedom and Resentment’ echoes the programmatic statement in the introduction of *Individuals*. Descriptive metaphysics, Strawson there explains, is concerned with the

massive core of human thinking which has no history—or none recorded in histories of thought; [...] categories and concepts which, in their most fundamental character, change not at all. (1959, p. 10)

As in ‘Freedom and Resentment’, in *Individuals*, Strawson specifies the subject of enquiry; also here, it is prompted by consideration of historical and cultural conceptual change (in particular, as to distance Strawson’s approach from the historicistic metaphysics of R. G. Collingwood (1940)).¹⁴⁵ This passage from *Individuals* also includes a further idea: that this “core of human

¹⁴³ It might be that this ambiguity of ‘we’—between a contrastive and an inclusive sense—is what encourages such mistaken construals of the Inescapability Claim as being a claim about us “as we are now” (Watson, 2014, p. 25). To this remark it should be added that Watson is clearer on the present matter than most (however, see also, Hieronymi, 2020; Russell, 2024). The mischaracterisation here remarked upon is specifically of the Inescapability Claim, and as such it is indeed a mischaracterisation. But Watson (2014, p. 30) also understands Strawson’s view as being that the participant stance which is given with the fact of human society is “the generic stance”, not any such stance characterized by specific, or specific versions of, the reactive attitudes.

¹⁴⁴ He admits of this as a shortcoming, though not a fatal one, in his response to Chattopadhyaya (1995): “Any such attempt [i.e., to describe the core of human thinking which has no history], as the history of philosophy shows, is likely to be affected by the state of knowledge and the general climate of opinion at the time at which the attempt is made. [...] But the attempt itself is worth making all the same.” (Strawson, 1995, p. 410).

¹⁴⁵ It also functions, if less directly, to distance descriptive metaphysics from conceptual analysis.

thinking [...] *has no history*”—this is an important point, which we will return to. Otherwise, the passage manifests the same kind of move as that towards the end of ‘Freedom and Resentment’.

We find this same move in passages throughout Strawson’s *oeuvre*, invariably in response to considerations of socio-historical variation between human conceptual schemes. Here, for example, in *The Bounds of Sense*:

The set of ideas, or schemes of thought, employed by human beings reflect, of course, their nature, their needs and their situation. They are not static schemes, but allow of that indefinite refinement, correction, and extension which accompany the advance of science and the development of social forms. At the stage of conceptual self-consciousness which is philosophical reflection, people may, among other things, conceive of variations in the character of their own situations and needs and discuss intelligibly the ways in which their schemes of thought might be adapted to such variations. But it is no matter for wonder if conceivable variations are intelligible only as variations within a certain fundamental general framework of ideas, if further developments are conceivable only as developments of, or from, a certain general basis. (2019, p. 34)

Also in *Scepticism and Naturalism*, following the very same dialectic as in *Individuals*,¹⁴⁶ we find Strawson making this move:

The human world-picture is of course subject to change. But it remains a *human world-picture* [...] So much of a constant conception, of what, in Wittgenstein’s phrase, is ‘not subject to alteration or only an imperceptible one,’ is given along with the very idea of historical alteration in the human world-view. (2008c, p. 21, original emphasis)¹⁴⁷

The central point of each passage is to focus enquiry on the *fundamental* features of *any* conceptual scheme. What Strawson is at each of these points concerned with elucidating are only the most *basic* features of any conceptual scheme, and these features as they figure in that role on any human conceptual scheme. Strawson understands conceptual variation, the form that different conceptual schemes take, as “developments of, or from, a certain general basis”. This general basis is what is given simply by the fact that any

¹⁴⁶ Again, the contrast is to Collingwood (1940).

¹⁴⁷ Strawson repeatedly stresses this line. More examples include Strawson’s (1995, p. 414) reply to Putnam (1995) and Strawson’s (2008c, p. 13) discussion of a distinction implicit in Wittgenstein’s metaphor of the river-bank. Part of the exchange between Martin Davies and Strawson also concerns this distinction (Strawson et al., 2008, p. 94f.).

conceptual scheme we can make intelligible to ourselves as such is a *human* conceptual scheme. It is *this* general basis, *as a general basis*, that is inescapable.¹⁴⁸

That this move is not only at the end of ‘Freedom and Resentment’ but recurrent throughout Strawson’s work supports the interpretative suggestion that the Inescapability Claim is best understood as an instance of a larger project. More precisely, it supports the reading that the Inescapability Claim is an expression of Strawson’s descriptive metaphysics.

So understood, what we see is that the Inescapability Claim is a claim about *only* that “fundamental general framework of ideas” that is “given along with the very idea of historical alteration in the human world-view”. It is only these concepts and categories, as they figure in that role, that are said to be “humanly necessary” (Strawson, 2008d, p. 42).

If we put the Inescapability Claim in terms of the *concept* of responsibility, we should say that *what* is inescapable is the concept of responsibility *in its non-historical form*.¹⁴⁹ Admittedly, what “non-historical form” means is not altogether clear. But we should at least say, in these terms, that what is *not* inescapable, in the sense of concern, is any particular *conception* of responsibility. That is, the form the concept of responsibility has in any particular conceptual scheme of some socio-historically situated group of concept-users is not the form in which it is inescapable. We may mark the distinction between the non-historical form of the concept and any historical form of that concept, as I have just done, by distinguishing between *the concept* of responsibility, on the one hand, and *a conception* of responsibility, on the

¹⁴⁸ Glock characterises Strawson (with Kant) as a “conceptual absolutist”, one who holds that “there could not be alternatives or changes to our conceptual scheme” (2009, p. 658). The point here is that this is true *only* with respect to the fundamental features of our conceptual scheme: Strawson is a ‘conceptual absolutist’ only with respect to *some* concepts or categories and, furthermore, only with respect to these concepts and categories *in their non-historical form*.

¹⁴⁹ While one of Strawson’s distinctive contributions to the philosophy of free will and responsibility indeed is his focus on the *practices* of responsibility and the *moral sentiments* or *reactive attitudes* (e.g., McGeer, 2019, p. 301; Russell, 2017b, p. 68; D. Shoemaker, 2020), this in no way renders talk in terms of the concept and conception of responsibility interpretatively unfaithful. In ‘Freedom and Resentment’, Strawson does talk not only in terms of the practice or the reactive attitudes, but also in terms of the concept of responsibility (2008a, pp. 1, 22, 23). And, while spelling the Inescapability Claim out in terms of inescapability of the reactive attitudes or the impossibility of universal objectivity of attitude are perhaps more common, to do so in terms of the *concept* of responsibility is not unprecedented (e.g., Alvarez, 2021; Bengtson, 2019; De Mesel & Heyndels, 2019). With respect to this issue, we may also remind ourselves that, according to Strawson, the concept is to be understood “from its use, from our practice” (1980, p. 265; also, 2011a, p. 90).

other.¹⁵⁰ So formulated, the Minimal Optimist claim is that *the concept* of responsibility is inescapable, while no particular *conception* of responsibility is.¹⁵¹

However, since any concrete ‘we’ is by definition *in history*, no concrete ‘we’ has the concept of responsibility purely in its *non-historical* form. It follows from this that an inescapability nevertheless remains at the level of the concrete ‘we’. Since we inescapably have the concept of responsibility, but cannot have the concept of responsibility *only* in its non-historical form, to have the concept of responsibility, as we necessarily do, we must have a conception of responsibility. To have *some* conception is thus also, in this extended sense, inescapable.¹⁵²

¹⁵⁰ There are some limitations of this idiom; we’ll touch on some of them below. Some misunderstandings, however, are best pre-empted. The *locus classicus* of the concept-conception is perhaps John Rawls: “Men disagree about which principles should define the basic terms of their association. Yet we may still say, despite this disagreement, that they each have a conception of justice” (1971, p. 5). Rawls (1971, p. 5, n. 1) explicitly follows Strawson’s close friend, fellow Oxfordian, and ‘Saturday Morning Meeting’-member, H. L. A. Hart (2012) here. This use of the distinction is different from our present use. We may see it as a difference of levels of explanation: where we are concerned with the conceptions of whole cultures in relation to some basic schema, foundation, or core that they all share, Rawls is concerned with the conceptions of individuals in relation to a culture that they are all taken to share. At the less general level at which Rawls employs the distinction, the distinction is typically cherished for facilitating intelligibility in disagreement (see e.g., Ezcurdia, 1998); we should recognise a similar need, and point, also at the more general level. Another difference is that, on a common usage of the distinction, it facilitates talk of *concept-mastering*, where this means assessing (individual or individuals’) conceptions as better or worse relative to how well they reflect the concept they are of (cf., Lalumera, 2014, p. 75ff.). This is *not* how the distinction is used here. Every conception unqualifiedly has or expresses the concept to equally well and to the same degree.

¹⁵¹ Already at this point, even if we were to keep understanding the Inescapability Claim as stating a psychological necessity (which we should not, as will become even clearer), the empirical plausibility criticism would at least be significantly attenuated. Something like this seems to be what Campbell is getting when he says that while we do not “have *specific abilities* to alter our hinge commitments”, such as that to responsibility, “it is likely that we have *general abilities* to alter our hinge commitments over time [...] as many free will skeptics admit” (2017, p. 47).

¹⁵² A clarificatory note: a ‘conception’, as we’re using the term here, is not *of* the concept. The concept of responsibility and any conception of responsibility are *of* the same thing, viz. responsibility. The concept is an abstraction from particular conceptions, allowing us to see these conception as of the same thing. Therefore, the cognitive or linguistic sophistication required to (implicitly or explicitly) have the concept of responsibility is the same as that required to have a conception of responsibility. I thank Niklas Dahl for prompting me to clarify this.

This already brings out an important weakness of talking, as I proposed, in the simplifying terms of ‘the concept’ of responsibility and various ‘conceptions’ of responsibility. Any conception of responsibility is of course embedded in a scheme of other ideas, some very closely related to responsibility, and some so related only on a particular conception, but not so on others. We talk generally of a conception, where this is understood as an outlook or framework of ideas; where it is not taken to be *of* any particular thing or anything in particular. When we talk of ‘conceptions of responsibility’, they must be understood as themselves being part of (being formed by and helping to form) a conception in this general sense of the term.

Talking freely of a pluralism of conceptual schemes may raise some eyebrows. Some scepticism is around with respect to talk of variations of conceptual schemes and indeed even of the very notion of a ‘conceptual scheme’, owing in part to Donald Davidson’s criticism in “One the Very Idea of a Conceptual Scheme” (1973-1974). The literature on this is vast. It will not be rehearsed in any summary form here. What we should however note is, first of all, that even if such scepticism were sound, it would not invalidate such talk in the interest of offering an interpretation of Strawson’s view. Strawson, as we’ve seen, talks freely of conceptual schemes and of the possible variety of different conceptual schemes. And, at least with respect to Davidson’s criticism, the scepticism of the very idea of a conceptual scheme is based, in part, on another scepticism—viz., about the idea of analyticity (cf., Quine, 1951). This scepticism, however, was famously rebutted by Strawson and Grice (1956).

We may furthermore note that, even if Davidson’s scepticism of the scheme/content distinction as a distinction between, on the one hand, an organising scheme and, on the other, an uninterpreted content is warranted—and Strawson (2019) also finds that understanding unacceptable—the distinction can still be upheld as an *intra-linguistic* distinction, between propositions that constitute our concepts, on the one hand, and propositions that employ these concept in making factual statements, on the other, like Wittgenstein’s (2009a) distinction between ‘grammatical’ and ‘empirical propositions’ (Glock, 2007, 2009).

In allowing for a variety of conceptual schemes, as of conceptions of responsibility, we might already have gone some way towards shedding the thought that there is something problematically conservative about Strawson’s dismissal of responsibility scepticism (cf. McKenna & Russell, 2008a, p. 10). While we necessarily have some conception, no particular conception, such as our current one, is inescapable. So, on the Minimal Optimist Reading, while

the Inescapability Claim holds in virtue of human nature, it is clear that, in this respect, human nature *underdetermines* our particular conceptual scheme.

This section was meant to clarify *what* is inescapable, according to Strawson. Something has surely been said on this account. However, on the understanding of the Inescapability Claim here proposed, our original question ‘What is inescapable?’ has in effect been reformulated to ‘What is *the concept* of responsibility, *in its non-historical form*?’ And *this* question has not yet been answered. Some may indeed contend, not unreasonably, that the idea of a ‘non-historical form’ is a terribly lofty notion and, further, that we do not really have any coherent grasp of ‘the concept’ in this sense but have (at most) a grasp only of a, or a few, conceptions. Without an answer to this further question, it may seem like we have only progressed in confusion, not in comprehension.

It was never Strawson’s project to say what, as we’re using the term, the concept of responsibility is. The interpretative project seems to come to something of a halt here. But we may plausibly carry this project a bit further even without such an answer. For the proposed delimitation of the Inescapability Claim to be illuminating, it is not necessary that we have an account of what the non-historical form of the concept of responsibility is. We do, in fact, have *some* grasp of the concept in the sense of concern. We in any case have enough of a grasp to sustain the distinction itself, between the concept and various conceptions, without any further criteria having been specified for what counts as a conception *of responsibility*. That we have a grasp of conceptions of responsibility other than our own already presupposes that we have some grasp of *the concept* of responsibility as distinct from any conception.

The question ‘What is the concept of responsibility, in its non-historical form?’ is indeed important for what anti-sceptical implications the Inescapability Claim, as here understood, has. The issue may be said to hinge on when to say that some conception of responsibility is in fact employed and when some conception is *not* a conception of responsibility, but something else. It would be a philosopher’s mistake, however, to assume that there has to be antecedently determinable criteria by which we can properly adjudicate any and every case of this sort.¹⁵³

¹⁵³ Plausibly, rather, this is a point at which philosophy needs history (cf., B. Williams, 2006b, 2014). In the free will debate today, however, the notion of ‘basic desert’ is meant to do something like this job—i.e., of supplying a principled criterion for distinguishing the, for the free will debate, relevant conception of responsibility from any, in that respect, irrelevant, alternative conception. On the notion of basic desert and its relevance, see for example McKenna (2019a, 2019b, 2019c), Menges (2023), Nelkin (2019), Pereboom (2019), and the paper collected in Sie and Pereboom (2015); for a brief overview, see Caruso (2021, sec. 1).

In the next chapter, more will be said about what the concept, already in its non-historical form, is; indeed, what it must be. I hope to show that it is essential for any actual conception of responsibility that it is, as it is sometimes put, a *backward-looking* conception of responsibility. In excluding instrumentalist (or *forward-looking*) conceptions of responsibility, while it does not straightforwardly amount to compatibilism, it does exclude what free will sceptics (e.g., Pereboom, 2014, 2021a) typically propose that we adopt instead of the supposedly problematic conception.

V. The Inescapability of the Bounds of Sense

Strawson's Analytic Kantianism

The Standard Reading understands the Inescapability Claim as a psychological thesis; an empirical claim, asserting a psychological necessity. This understanding of the claim is unfounded and positively mistaken. The Inescapability Claim is better understood if seen as an expression of Strawson's descriptive metaphysics. However, more needs to be said in order to replace the Standard Reading.

In this section, we'll begin by asking what *status* the Inescapability Claim is supposed to have, primarily in order to see what *kind* of claim it is (if not an empirical claim). The suggestion is that we consider it as stating a framework commitment, conceptual but not analytic in character. We then turn to the issue of elucidating the *sense* in which the concept of responsibility is said to be inescapable or necessary (if this is not a psychological necessity). I'll propose that, in light of the kind of claim the Inescapability Claim is, there are *two* different senses in which the concept of responsibility is held to be inescapable: one is a *practical necessity*, the other a *conceptual necessity*.

While the Standard Reading draws on Strawson's references to Hume, we probably do better if we draw, rather, on the references (often co-occurring) to Wittgenstein. The reading developed here is of that sort.¹⁵⁴ However, it may be contended that this is only a reflection of the interpreter's bias, a matter of placing emphasis with one's sympathies. It is not.¹⁵⁵ But perhaps the issue is

¹⁵⁴ Readings of this sort include Campbell (2017), De Mesel (2018), Heyndels (2020), and Rummens & De Mesel (2023).

¹⁵⁵ The Wittgensteinian influence on Strawson is broad and undeniable; we have seen this throughout this thesis. Beside those cited in the previous footnote, in connection to Strawson

best settled indirectly. Rather than making it an issue of where to place the emphasis, on Hume or Wittgenstein, we might be able to work our way up to a plausible understanding of the Inescapability Claim by a different path: by considering Strawson's analytic Kantianism, and thus further elaborating on the fact that the Inescapability Claim is expressive of Strawson's descriptive metaphysics.¹⁵⁶

Having enlisted Kant as a descriptive metaphysician when introducing this project in *Individuals*,¹⁵⁷ Strawson turned directly to Kant in *the Bounds of Sense*, attempting to exorcise what is unacceptable in his system from the real insights. What he is primarily out to save is the idea, broadly speaking, that there is a lower and an upper limit to sense:

on the one hand, that a certain minimum structure is essential to any conception of experience which we can make truly intelligible to ourselves; on the other, that the attempt to extend beyond the limits of experience the use of structural concepts, or of any other concepts, leads only to claims empty of meaning. (Strawson, 2019, p. xix)

This is the framework Strawson's descriptive metaphysics seeks to elucidate: "the most general features of our conceptual structure" (Strawson, 1959, p. 9). To extend a concept beyond this framework is to extend it beyond the area which defines our competence with that concept: "So using it, we shall not merely be saying what we do not know; we shall not really know what we are saying" (Strawson, 2019, p. 5).

Strawson's central criticism of Kant is that he violates his own principles by trying "to draw the bounds of sense from a point outside them, a point which, if they are rightly drawn, cannot exist" (Strawson, 2019, p. xix). To explain

on responsibility, see Bengtson (2019, 2020), De Mesel (2018, 2022a, 2022b), De Mesel & Heyndels (2019). Something might seem a bit off with this dichotomy, since by Strawson's own lights there is "a profound community" (2008c, p. 15) between Wittgenstein and Hume with respect to the relevant issue. However, see Strawson's "Reply to Hilary Putnam" (1998d) for a telling discussion regarding his relation to Hume and to Wittgenstein and some interesting remarks on the relation between them, as Strawson sees it. Besides 'Freedom and Resentment', it is *Scepticism and Naturalism* that encourages a Humean emphasis. With respect to this, it is interesting to note the dialectical structure which Strawson gives the issues there discussed: Strawson's *Scepticism and Naturalism* takes as its point of departure Moore's 'Defence' (1925) and 'Proof' (1939), just as Wittgenstein's *On Certainty* does.

¹⁵⁶ On Strawson's analytic Kantianism, see in particular the collection *Strawson and Kant*. For further helpful discussion, see for example Gomes (2016, 2024), Glock (2007, 2012, 2024).

¹⁵⁷ A work "subtly and in part consciously influenced by" Kant's *Critique of Pure Reason* (Strawson, 2003, p. 8).

why the bounds of sense are so drawn, which is for Kant to account for “the mind producing Nature as we know it out of the unknowable reality of things as they are in themselves” (Strawson, 2019, p. 11), Kant engages in, as Strawson puts it, “the imaginary subject of transcendental psychology” (2019, p. 21). But this subject is indeed imaginary and cannot provide the account Kant needs it to provide.

In his discussion of Strawson’s criticism, Anil Gomes (2024) proposes not only that we see Strawson’s criticism as an instance of a very old challenge—of the need for a *metacritique*: the need to account for the grounds of those claims which make up the argument for what the bounds of sense are. He furthermore proposes that we pose this old challenge also to Strawson: What is the status of the claims that make up Strawson’s descriptive metaphysics? More importantly still, for our purposes, is to ask what the status of the conclusions are: What status does the description of the limiting structure of any conception of experience that we can make intelligible to ourselves as such have?

Against Kant, Strawson declares that no “high doctrine”, such as Kant’s doctrine of synthesis, is necessary here:

In order to set limits to coherent thinking, it is not necessary, as Kant, in spite of his disclaimers, attempted to do, to think both sides of those limits. It is enough to think up to them. (Strawson, 2019, p. 34)

Seeing that descriptive metaphysics does not differ from conceptual analysis “in kind of intention, but only in scope and generality” (Strawson, 1959, p. 9), a natural thought is that, in thinking “up to” the limits of coherent thinking, what we are doing is an extended form of conceptual analysis and, therefore, that the claims so arrived at are pieces of *analytic* knowledge.¹⁵⁸

Jonathan Bennett’s (1966, pp. 41-42) reading is along these lines. He suggested that, just as we should understand Kant’s own claims and conclusions to be ‘unobvious analytic truths’, we should understand Strawson’s claims and conclusions in *Individuals* as ‘unobvious analytic truths’. What is *unobviously analytic* is established by an exceedingly lengthy chain of analytic truths, which is why the truths so arrived at are unobvious yet analytic.

However, as Gomes (2024, p. 154) convincingly argues, this is not Strawson’s view. First, in his review of Bennett’s *Kant’s Analytic* (1966),

¹⁵⁸ Strawson’s *analytic* Kantianism consists, indeed, in taking the *Critique*’s central point to be “an *analysis* of complex connections between concepts such as experience, self-consciousness, objectivity, space, time and causation” (Glock, 2003a, p. 1).

Strawson (1968, p. 334) argues that the relevant claims in Kant’s system do not fit the model of unobviously analytic truths. Furthermore, in that same review, Strawson questions the very need for invoking ‘analyticity’ “to preserve the respectability of metaphysics” (1968, p. 335). Finally, Strawson thinks that “[w]hether or not we choose to entitle the propositions descriptive of [the fundamental general] structure [of any conception of experience such as we can make intelligible to ourselves] ‘synthetic *a priori*’, it is clear at least that *they have a distinctive character or status*” (Strawson, 2019, p. 33, emphasis added). At least the claims that describe the framework itself, then, are not, in Strawson’s view, merely analytic truths (whether obvious or unobvious such), but claims with a distinctive character or status.¹⁵⁹

What ‘distinctive character or status’ do they have, then? Gomes (2024, p. 156) answer is that they are *non-analytic but still conceptually necessary*. A clear reason to think this, he shows, is that Strawson himself, by his old student Quassim Cassam’s report, “sometimes described them, somewhat mysteriously, as non-analytically but still conceptually necessary conditions” (Cassam, 2016, p. 915). Assuming that this is right, it nevertheless remains a question how we are to understand *that* status.¹⁶⁰ Hacker (1996, p. 177; 2001,

¹⁵⁹ See also, for example, Strawson (1995, p. 415).

¹⁶⁰ Without saying further what status that is, Gomes (2024, pp. 157-163) offers an account of how the claims could be neither empirical nor analytic, yet conceptually necessary and known *a priori*. He gets at this via a possible similarity in the methods of G. E. Moore and Strawson. The central feature he takes them to share is the involvement of the first-personal perspective in conceptual analysis. I’ll not evaluate Gomes’s suggestion in any detail, but only raise two further, if minor, points that perhaps support Gomes’s account of Strawson’s method, and then one, all too brief, but not insignificant, reservation. First, in section IV of ‘Freedom and Resentment’, Strawson says that he will, for one thing, ask under what conditions the reactive attitudes are and are not appropriate, and further—this is the interesting bit—“what it would be like, what it *is* like, not to suffer them” (Strawson, 2008a, p. 7, original emphasis). This has that phenomenological tone that Gomes takes to be an implicit aspect of Strawson’s methodology—making claims based on a familiarity with the concept of (a particular) experience; claims that may seem empirical, but aren’t. (Strawson elsewhere describes what he was doing in ‘Freedom and Resentment’ as referring us “to the phenomenology of the moral life” (1998b, p. 262)). Second, Strawson remarks that “all fruitful philosophical enterprise must be” individual (1998a, p. 19), because “truth in philosophy” unlike in the special sciences, “is so complex and many-sided, so multi-faced, that any individual philosopher’s work, if it is to have any unity and coherence, must at best emphasize some aspect of the truth, to the neglect of others” (2005, p. viii). This provides, if nothing else, a further point for the involvement of, not quite the first-personal, but one’s individual perspective, as well as a further contrast, in virtue of this involvement, with the special sciences. Last, a reservation: Strawson’s language seems more geared to our *shared* experience, rather than the first-personal experience—the facts as *we* know them. (Something similar might perhaps be said also about Moore’s method (Vanrie, 2021).) The

p. 361) suggests that these conceptual connections express ‘norms of representation’, that they are part of ‘grammar’ in a Wittgensteinian sense (cf., Gomes, 2024, p. 157, n. 18). Perhaps this is too restrictive still (which of course depends on one’s understanding of ‘grammar’).¹⁶¹ Although interesting, we need not linger on this issue any further.

What we have gathered from this discussion so far, working with the assumption that the Inescapability Claim is an expression of Strawson’s descriptive metaphysics, is that the Inescapability Claim is not, contrary to the Standard Reading, an empirical claim. Rather, Strawson works from *within* our conceptual scheme *up to* its limits. While this does not mean that the laying down of these limiting features is (merely) a matter of demonstrating analytic truths, the claims established by descriptive metaphysics—that something is a limiting feature of any conceptual scheme we can make intelligible to ourselves as such—purports to have a distinctive character or status, capturing a (broadly speaking) conceptual necessity.

Perhaps their distinctive status is best explained, as Glock (2024, pp. 122, 142; also Peter M. S Hacker, 1996, p. 177) proposes, in terms just of them having that peculiar role in our conceptual scheme that they have.¹⁶² So understood, their distinctive character consists in that they are the commitments constitutive of that “minimum structure” (Strawson, 2019, p. xix). Let’s call them *framework commitments*.

use of ‘we’ here might indeed be that of an inviting ‘we’, but it still does not seem to quite fit the model that Gomes proposes; at least, the emphasis seems different.

¹⁶¹ Some of the claims may perhaps be understood as ‘hinge commitments’, and while some (e.g., Moyal-Sharrock, 2004a) take these to be part of grammar, others (e.g., Glock, 2009, pp. 657–667; Glock, 2016, pp. 288–292) take them to form a separate category.

¹⁶² Glock also suggests that one kind of conceptual connection involved in supporting these claims are “conceptual truths [that] are non-trivial because they are *not definitional* in even a catholic sense of that label” (2024, p. 143). Picking up on Bennett’s terminology, we may characterize such claims as *unobvious conceptual truths*. Such truths are established by making use of a *third concept* in establishing the conceptual connection; where this third concept is not itself analytically or grammatically related to the other relata. For example, consider how Strawson formulates the problem of elucidating the relations between *belief*, *desire* and *action*: “‘In men, or indeed in any rational being, the three elements of belief, valuation (or desire), and intentional action can be differentiated from each other; yet no one of these three elements can be properly understood, or even identified, except in relation to the others.’” (1992a, p. 80). A question remains: what relation holds between either of the unobviously conceptually connected concepts and that third concept, a relation that is conceptual, yet “*not definitional* in even a catholic sense”?

Framework Commitments

Characterising them as framework *commitments* is meant to capture their *non-rational* or *pre-rational* character. Strawson repeatedly uses the notions of ‘belief’ and of ‘proposition’ in describing what we are committed to, but it is not quite clear that these adequately capture the character or role of these limiting features of our conceptual scheme. Regarding this issue, Heyndels (2020, p. 100) argues that we should ascribe the following claim to Strawson:

[Not A Belief] The concept of belief (or proposition) is inadequate to characterize our natural commitment [...]¹⁶³

This interpretative suggestion is getting at an important truth—i.e., that being so committed is fundamentally a matter of acting in a certain way; that the commitment is non-rational or pre-rational. However, the attribution of [Not a Belief] to Strawson is nevertheless not straightforward.

Comparing our commitment to the participant stance with that to inductive reasoning, Strawson does describe the commitment(s) as “original, non-rational (not *irrational*), in no way something we choose or could give up” (2008a, n.7). But then again, he regularly speaks of them as “nonrational commitment[s] to *belief* [...]” (Strawson, 1998b, p. 260, emphasis added). Elsewhere he remarks on an aversion in both Hume and Wittgenstein to using ‘belief’ or ‘proposition’ in connection with these framework commitments (Strawson, 1998c; 2008c, p. 14). Perhaps his occasional use of scare quotes reveals some hesitancy on his own part concerning these terms (Strawson, 1998c, p. 370; 1998d, p. 288). But he also makes very clear that we do not need follow Hume and Wittgenstein in, as he puts it, their “extravagant conclusions” (Strawson, 1998c, p. 371; also, Strawson, 2008c, p. 13f.). While Wittgenstein (also in *On Certainty* (1969, §§35, 36)) took such metaphysical claims as ‘There are physical objects’ to be nonsense, Strawson does not. Then again, that he himself allows them to be characterised as “propositions or *crypto-propositions*” (2008c, p. 18, emphasis added) may give the impression of some irresolution or qualification on this issue.¹⁶⁴

¹⁶³ See also Glock (2022, pp. 448, 450).

¹⁶⁴ Perhaps we should see Strawson as closer to Collingwood than to Wittgenstein on this matter. Collingwood’s notions of *absolute presuppositions* and *absolute proposition* might be a helpful in spelling out Strawson’s notion of framework commitments. An absolute proposition is of course propositional, but it lacks truth-value. When it figures as an absolute presupposition, it is not the apparent semantic content that is interesting, but its expression in the form of a particular practice.

In unpacking Strawson's position here, we perhaps do best to understand our natural commitment as being—indeed not, in the first instance, to some belief, but—to a *conceptual structure*. This is a point at which the idiom of the concept of responsibility, rather than that of the reactive attitudes or our practices, awards perspicuity. Focusing on the conceptual, we can distinguish two different aspects of, or sides to, Strawson's use of framework commitments—one up-stream and the other down-stream from the conceptual structure itself.

The first is that the *source* of these commitments is *nature*, our human *form of life*. We are not so committed because we have decided to be (e.g., Strawson, 2008a, pp. 14, 20; 2008c, p. 14; Strawson et al., 2008, p. 92f.). We do not hold these commitments *for reasons*, for they are “outside our critical and rational competence in the sense that they define, or help to define, the area in which that competence is exercised” (Strawson, 2008c). (Here, Bennett's (1980) and Watson's (2008) ‘non-propositionalism’ emerges in its proper guise.) In this respect, they are non- or pre-rational. These “unavoidable natural convictions, commitments, or prejudices” stem (on Hume's simpler account) from Nature and (on Wittgenstein's more complicated account) from social practices (Strawson, 2008c, p. 14f.). They are in the first instance reflections of our practice, our form of life, our human nature.

The second aspect of our framework commitments, down-stream from the conceptual structure that they are commitments to, is that in having certain concepts, we in effect have certain beliefs. In effect, therefore, these are natural commitments to belief. From the claim that we necessarily have certain concepts and that we at least need to regard ourselves as being able to apply these concepts, Strawson immediately moves to the claim that we necessarily believe, for example, that people in general are responsible.¹⁶⁵ The first part of this move is not our present concern.¹⁶⁶ The second part of this move presupposes the view that belief-formation follows naturally from having concepts, in the sense that whether we recognise something we encounter as a thing *of that sort*, as falling under some concept, is not something normally up

¹⁶⁵ For one compact instance of this, see Strawson et. al. (2008, p. 91f.).

¹⁶⁶ It involves the idea that in having a concept, we have at least an implicit mastery of the use of a certain range of expressions employing that concept (Strawson, 1992a, p. 6f.). If it is to show the sceptical challenge unintelligible, it also needs to show establish certain conceptual connections: that in having some concepts and regarding ourselves as able to apply these, we also need to have other concepts, and regard ourselves as able to apply them. This is the key to developing transcendental arguments, but not our present issue.

to the will (Glock, 2024).¹⁶⁷ While one may not act on it in a given case, if we have, for example, the concept of responsibility, we cannot but believe that someone *is* responsible when the criteria for that concept are taken to apply. We necessarily regard them as responsible under such conditions. This does not mean that people in fact are responsible, but it does mean that we believe that they are if we take the concept to apply.

If we postulate a view like this, then we may account for Strawson's move from necessarily having certain concepts (in virtue of our human form of life), to us necessarily having certain *beliefs*, and thus explain the characterisation of these commitments as "nonrational commitment[s] to belief" (Strawson, 1998b, p. 260). To understand Strawson's framework commitments, the proposal is, we are helped by focusing on the framework—the conceptual structure—that they are commitments to; this connects the non-rational *source* of our belief-system with the *content* of that belief-system.

Doubly Inescapable: Two Senses of Necessity

Turning now to the final and key issue of this section: In what *sense* is the concept of responsibility inescapable? I think there are two different senses in which the concept of responsibility is inescapable, according to Strawson. Our framework commitments are both *practically* inescapable and *conceptually* inescapable. This is reflected in their character as framework commitments—that they are *pre-rational commitments* and commitments to a *conceptual framework*.

In chapter 2, we considered *Glock's Question*: What is the nature of the sceptic's mistaken, according to Strawson? The impetus of this question was a real confusion regarding how to interpret Strawson's anti-sceptical approach. While defusing the apparent conflict between different ways Strawson characterises the sceptic's mistake, I stressed the *senselessness* of the demand for a justification of our conceptual scheme. As we may put the point now, the emphasis was placed on the senselessness of attempting to transgress the bounds of sense. With respect to our present question, the impossibility of either doubt or justification without presupposing the conceptual framework

¹⁶⁷ This is of course not to deny that we may actively seek to see things differently, as in Iris Murdoch's (2001) famous example of the mother (M) and her daughter-in-law (D), by attending to *other* features of the object, focusing on what other *sorts of thing* it is. Nor does this view preclude actively refusing to *use* some concept—for example, as Oscar Wilde is said to have done when, invited by a cross-examiner to concede that some piece of conduct of his was blasphemous, he responded, "Blasphemous is not a word of mine".

itself amounts to one kind of inescapability or necessity: a *conceptual necessity*. But there is also another sense of inescapability or necessity—a *practical necessity*—operative in Strawson’s approach. Indeed, as we’ll soon see, this further sense of inescapability is crucial in the case of responsibility.

One kind of inescapability is a *practical inescapability*: it follows from *the necessity of the practice* of which our conceptual scheme is expressive. If there is to be anything recognisable as a human form of life, there will be human practices. These practices are expressive of certain needs and situations. The conceptual scheme “employed by human beings reflect, of course, their nature, their needs and their situation” (Strawson, 2019, p. 34). These needs and situations are the “natural foundations of our logical, conceptual apparatus” (Strawson, 2011a, p. 86). According to the Inescapability Claim, our conceptual scheme stems from these needs and situations. It is in virtue of these that a *certain* minimum structure, which included the concept of responsibility, is practically necessary for us.

At this point, we might think that the Standard Reading regains its footing. Admittedly, human psychology is relevant. But it remains a fact that an explanation of human nature in exclusively psychological terms is seriously inadequate. Perhaps even this admittance grants too much to empirical psychology, or to any more complicated story of the kind.¹⁶⁸ Glock (2012, pp. 293–297; 2017, n. 1; 2022, p. 450; 2024, p. 123) calls the study Strawson presents, of the natural underpinnings of our conceptual scheme, “explanatory metaphysics”, and that term—‘metaphysics’—seems perfectly apt. In the next chapter, I offer what is, I argue, a kind of explanatory metaphysics in the sense that Strawson outlined. It is not an evolutionary account, but it is a kind of genealogical account. The sense of inescapability pertaining to the practice according to that explanation is a *practical necessity*. We’ll then see the merits, interpretative and otherwise, of reading Strawson as implicitly relying on something like this.

The second sense of inescapability is *the inescapability of our conceptual structure*. The Inescapability Claim, understood as descriptive of the bounds

¹⁶⁸ “An exponent of a [...] thoroughgoing naturalism could accept the question, What causes induce us to believe in the existence of body? as one we may well ask, as one that can be referred to empirical psychology, to the study of infantile development; but would do so in the justified expectation that answers to it would in fact take for granted the existence of body” (Strawson, 2008c, p. 10). Then again, at one point, Strawson himself seems to suggest that an evolutionary account may explain why we have the concepts that we have (Strawson, 1998d, p. 289). But this is probably not how to understand his here. It is at odds with how he presents the relevant kind of explanation elsewhere. We’ll consider this in more detail in the next chapter.

of sense, is expressive of a non-analytic yet conceptual necessity. As a framework commitment constitutive of that “minimum structure”, doubt with respect to it is tantamount to doubt of the general framework itself. To attempt to doubt these framework commitments—or, for that matter, to attempt to justify them—“is simply to show a total misunderstanding of the role they actually play in our belief-system” (Strawson, 2008c, p. 15). Such doubt attempts to doubt “the general framework within which all doubts and queries and reasoning makes sense”—and, without which, they do not (Strawson et al., 2008, p. 92). For this reason, to attempt to doubt the general framework itself “doesn’t make sense for us” and “it’s something inescapable in that sense” (Strawson et al., 2008, p. 92). There is this sense of inescapability on Strawson’s account that is first of all conceptual, in that ecumenical sense which we have seen that Strawson takes there to be. What is at issue here, in the first instance, is a matter of intelligibility.¹⁶⁹

Throughout this chapter, the Inescapability Claim has been treated as just one among other claims that seek to lay bare, as Strawson puts it, “a certain minimum structure [...] essential to any conception of experience which we can make truly intelligible to ourselves” (2019, p. xix). In a number of respects, I hope to have shown this to be an illuminating parallel. But now we need to recognise “a twist or complication” in the case of responsibility (Strawson, 2008c, p. 31).¹⁷⁰

¹⁶⁹ “In the first instance” because the intelligibility of concern may ultimately be understood as a practical failure, as some Wittgensteinians would argue; it is not implausible that we are to count Strawson among them. An indication of this, an amusing one at that, is Strawson’s reply to Wesley Salmon’s (1957) accusation that by taking human nature as the non-rational foundation of our inductive reasoning, Strawson makes it an arbitrary choice whether to reason inductively. To this, Strawson remarks, “Suppose I am convinced that there is nothing to choose, as far as Reason goes, between the ‘basic canons’ of induction, and a consistent counter-inductive policy. Is an ‘arbitrary choice’ then really open to me? Is it? (Just try to make it.)” (1958, p. 21) This is, but is not only, a quick and sarcastic retort of a younger Strawson. Also in *Scepticism and Naturalism*, he specifies what he means by entertaining *serious* doubt as follows “(Serious = actually making a difference)” (2008c, p. 42); this strikes a familiar tone.

¹⁷⁰ Strawson does not say “in the case of responsibility” but “in the moral case”. In *Scepticism and Naturalism*, there is not that distinction, on which ‘Freedom and Resentment’ pivots, between personal and moral reactive attitudes. In light of that distinction, I think “in the case of responsibility” better captures what is the issue.

VI. A Twist or Complication

The concept of responsibility is *not* an essential feature of *any* conception of experience which we can make truly intelligible to ourselves. We do have a conception of experience which we can make intelligible to ourselves which is not defined by a commitment to the concept of responsibility—the detached perspective, which we sometimes take toward other people. That is the “twist or complication” in the case of responsibility. It is in this respect disanalogous to those other framework commitments we have compared it to.

The concept of responsibility is an essential feature of *one kind of experience*. It is conceptually necessary for the participant stance, for being involved in interpersonal relationships. This fact puts pressure on that other aspect of the Inescapability Claim: the *practical necessity* of the concept; or, which comes to the same thing, the practical necessity of *that* kind of experience.

When we delimited what is inescapable above, the following passage from ‘Freedom and Resentment’ served as our key; especially the first emphasis:

In the absence of *any* forms of these attitudes it is doubtful whether we should have anything that *we* could find intelligible as a system of human relationships, as human society. (Strawson, 2008a, p. 26, original emphasis).

Note now the second emphasis in this passage. What the emphasis on ‘we’ signifies is the *inescapability of the bounds of sense*: without a conception of responsibility, we cannot make anything intelligible to ourselves as a system of human relationships, as human society.

But this inescapability is in one respect insufficient. In virtue of the twist in the case of responsibility, the Inescapability Claim receives a conditional character: *If* we are to have anything we can make intelligible to ourselves as human society, *then* we need to have some conception of responsibility. And our question becomes: Do we need to have anything of that sort? Could we not always take the objective attitude?

This issue puts pressure on the *practical necessity* of that experience for which responsibility is conceptually necessary. That it does, however, is not generally recognised. Gary Watson’s (2014) reaction to the twist or complication in the case of responsibility is a different one. He too holds that the inescapability should be relativised in light of this twist, but he does not do so in a way that puts pressure on the practical inescapability. According to Watson, the “relative inescapability” (2014, p. 23) Strawson propounds is an

inescapability relative to our “sociality”, (2014, *passim*).¹⁷¹ What the twist or complication in the case of responsibility means, according to Watson, is only that we might not have “the evaluative orientation on which our competence with the concept of responsibility depends” (2014, p. 23f.). But this is no real threat, according to Watson, because “to imagine escaping the participant stance altogether is to imagine escaping one’s ‘own humanity’” (2014, p. 23).¹⁷² While this more or less gets at, what I’ve called, the conceptual necessity,¹⁷³ it does not appreciate what is special about the case of responsibility. This conceptual necessity holds equally for the other framework commitments; appealing to it cannot explain what is disanalogous with the case of responsibility. At most, it emphasises that only those that *de facto* are committed to the interpersonal attitudes are inescapably committed to the concept of responsibility. But this seems insufficient in light of the fact that we also have at our disposal the detached perspective. For the question now is: Should this perspective dominate our lives?

According to Strawson, the commitment to responsibility is “a condition *of* our humanity” (Strawson, 2008c, p. 26, emphasis added)—it is not merely a condition *on* “our humanity”. While the conceptual necessity holds for our ‘sociality’, the practical necessity holds with respect to ‘society’. And that is part of the claim: “The existence of the general framework of attitude itself is something we are given with the fact of human society” (Strawson, 2008a, p. 25). The detached perspective, while we have recourse to it on some occasions, *cannot* dominate our lives.

While the concept of responsibility is only an essential feature of one kind of experience—which already implies the relevant sense in which the commitment is, in Watson’s terms, a condition *on* “our humanity”—it is also the case that, according to Strawson, *this* kind of experience is itself given with our human form of life. If we are to displace the possibility of universal objectivity of attitude, this further point is necessary. If we are to address

¹⁷¹ Possibly, Watson’s construal in terms of ‘sociality’ rather than ‘society’ is also an instance of the individualistic assumption.

¹⁷² The quote is from Strawson, and worth taking in full: “what is above all interesting is the tension there is, in us, between the participant attitude and the objective attitude. One is tempted to say: between our humanity and our intelligence. But to say this would be to distort both notions” (2008a, p. 10). Watson, it seems, exploits this distortion.

¹⁷³ Watson’s claim gets at it *only if* “our humanity” is not taken to mean what Strawson is “tempted to say”, but means something less moralistic. The reason why this is necessary in articulating Strawson’s position is just the fact that it is a *Minimal Optimism*, that his claim concerns that essential “minimal structure” and not any particular conception of, *inter alia*, responsibility.

scepticism about responsibility in particular, that is, this point is all but irrelevant (pace, McGeer, 2014; Sars, 2022a; Watson, 2004).

In the next chapter, I'll offer an explanation that shows that that the concept of responsibility is indeed practically necessary for human society, thus substantiating this part of Strawson's claim. If this is right, then we cannot opt out—even if we are willing to pay the price of “our humanity” or “sociality”.

The ‘we’ of the last sentence is ambiguous. Clarifying it allows us to address another misconception of the Inescapability Claim which we remarked on before. What we called ‘the individualistic assumption’ is that the ‘we’ here means that *no single individual* (i.e., no one of us) can abandon the commitment to responsibility. But this is plausibly not what is necessary for human society. What is necessary for the general framework to be in place, according to Strawson, is a certain *regularity* in practice: “it is a condition of the existence of any form of social organization, of any human community, that certain expectations of behaviour on the part of its members should be pretty regularly fulfilled” (2008d, p. 33).¹⁷⁴ There is no reason to assume that the maintenance of this regularity in practice is distributed in accordance with an egalitarian principle. An individual lacking the framework commitment to responsibility, unable to (sincerely) share in our interpersonal relationships, is not (*pace* the empirical plausibility criticism) rendered inconceivable on this construal of the Inescapability Claim.¹⁷⁵

VII. Conclusion

The ambition of this chapter has been to replace the Standard Reading of the Inescapability Claim. The reading here proposed pivots on the point that the Inescapability Claim should be seen as an instance of Strawson's project of supplying a descriptive metaphysics of that core of human thought which, in its fundamental character, has no history. In several respects, the Inescapability Claim should be seen as paralleling other claims of descriptive metaphysics.

Understood as a claim of descriptive metaphysics, the Inescapability Claim should neither be seen as an empirical claim (about human psychology), nor

¹⁷⁴ This is a familiar Wittgensteinian point (e.g., 2009a, §242).

¹⁷⁵ Similarly, Williams (1985, p. 28) affirms the impossibility of us doing without ethical life, while allowing that any individual agent may nevertheless live outside it. Watson (2014, p. 23) makes what is at least a similar point, that some people in fact lack ‘a moral sense’ and hence stand outside that general framework which is defined by a commitment to responsibility.

as stating an analytic truth (e.g., that ‘being responsible’ holds in virtue of the meaning of ‘person’), but rather as laying bare a framework commitment. That we have a framework commitment to responsibility means that, in virtue of our human form of life, we are necessarily committed to the concept of responsibility. The concept of responsibility is a feature already of this minimum structure, but this also entails that we are not inescapably committed to the concept of responsibility as it figures in some socio-historically local conceptual scheme (i.e., we are not inescapably committed to any particular conception of responsibility). The commitment to this conceptual framework is pre-rational—not something we chose or hold for reasons. As we do not hold framework commitments for reasons, the issue of whether we are justified in holding them is senseless, showing a misunderstanding of the role these commitments has with respect to our practices of giving reasons. Any justification, as any doubt, presupposes the conceptual framework itself.

However, according to Strawson, the case of responsibility is not insignificantly disanalogous to the other framework commitments. There is here, as he puts it, a “twist or complication”. This, I argued, puts pressure on the practical necessity of the concept of responsibility. The next chapter aims to supplement Strawson’s approach with a kind of explanation that brings out the practical necessity of this concept.

The Inescapability Claim, as here understood, was said to provide the route through which we may make good on Strawson’s Promise. This chapter has first of all aimed to replace the Standard Reading. The basic sense in which we might say that we have made good on Strawson’s Promise is that, for the Minimal Optimist, is it not a question *whether* we are responsible. We cannot but see ourselves as responsible. At this point, still, this might seem merely conservative. Admittedly, as of yet, it has not been shown how exactly the Inescapability Claim may ground this attitude. It is the aim of the following chapters to substantiate the claim and show that, and in what further respects, it may be said to make good on Strawson’s Promise.

6. The Unanswered Question

It is important to recognize the diversity of possible systems of moral demands [...] But it is also important to recognize that certain human interests are so fundamental and so general that they must be universally acknowledged in some form and to some degree in any conceivable moral community.

P. F. Strawson, 'Social Morality and Individual Ideal'

I. Introduction

In 'Freedom and Resentment', Strawson sets up a question for his enquiry which he explicitly puts to the side, and it has remained neglected. The object of this chapter is to supplement Strawson's approach with an answer to that heretofore unanswered question.

In section I and II of 'Freedom and Resentment', Strawson lays out the relevant dispute in his now well-known, idiosyncratic way. In section III, he seeks to attune us to some, as he calls them, "commonplaces" of human life—in particular, how much we mind, how we emotionally react to, people's attitudes and intentions towards other people; it is here he introduces the philosophical community to the 'reactive attitudes'.¹⁷⁶ Then, in section IV, with the very first lines, Strawson raises three questions about the reactive attitudes:

- (1) What are the general causes of these attitudes?
- (2) What are the particular conditions under which they do or do not seem natural or reasonable or appropriate?

¹⁷⁶ Or rather *reintroduces*: "It is a pity that talk of the moral sentiments has fallen out of favour. That phrase would be quite a good name for that network of human attitudes" (2008a, p. 25f.).

(3) What would it be like, what is it like, not to suffer them?¹⁷⁷

Strawson himself is “not much concerned with the first question”, but professedly more so with the other two. This interest has proven to be part of the legacy of Strawson. Articulating the appropriateness (or fittingness) conditions for the reactive attitudes in general or for (an ever-growing range of) particular reactive attitudes is the dominant focus in the literature. Investigations of the constitutive commitments and features, as well as the very phenomenology of, the participant stance and the objective attitude are likewise central concerns in the contemporary discussion.¹⁷⁸ In this light, the project of this chapter, to answer the first question, might not seem very Strawsonian. But an answer to this question is just what we need if we are to appreciate Strawson’s way with the issue of free will and responsibility, so I’ll argue. The here proposed answer to the unanswered question of ‘Freedom and Resentment’ will not be, only because it cannot be, Strawson’s answer—but it will be the answer Strawson *should* have given.¹⁷⁹

In the process, we will return to several issues that have been left underdeveloped. In particular: What is ‘the concept’ of responsibility? Can we make any further sense of it being a ‘non-historical’ concept? Is the concept of responsibility really *necessary* for us and, if so, why? Does the practical necessity of the concept show that determinism is fundamentally irrelevant for responsibility?

Section II and III anchor the need to answer the unanswered question in Strawson’s account, both in the sense that Strawson took such an answer to be needed for philosophical understanding and in the sense that such an answer becomes needful in light of the rejection of the demand for justification,

¹⁷⁷ The relevant passage in full: “It is one thing to ask about the general causes of these reactive attitudes I have alluded to; it is another to ask about the variations to which they are subject, the particular conditions in which they do or do not seem natural or reasonable or appropriate; and it is a third thing to ask what it would be like, what it is like, not to suffer them. I am not much concerned with the first question; but I am with the second; and perhaps even more with the third” (Strawson, 2008a, p. 7).

¹⁷⁸ ‘What would it be like not to suffer the reactive attitudes?’ is a central question for such sceptical positions as Pereboom’s (2001, 2014). For other kinds of answers, see e.g., Sommers (2007), and Hutchison (2018).

¹⁷⁹ This should not be taken to imply that interpretative standards are irrelevant; the project is still very much a *reconstruction* of Strawson’s view, if a *creative* reconstruction. Since this aspect of Strawson’s approach is intentionally left out and there is no ‘text’ to count as Strawson’s position on this issue, the answer here developed cannot be said to be an interpretation of Strawson’s position on this issue.

something Strawson seems to recognise. Section IV presents the answer here proposed to the unanswered question of ‘Freedom and Resentment’—a pragmatic genealogy. We consider an apt example of such an explanation, of blame, drawing on some key remarks by Strawson and, if in a more *ad hoc* manner, of some further features of the concept of responsibility. Section V and VI makes the case, in light of what has been said so far, why the answer here proposed is the answer Strawson should have given. Section VII makes explicit the senses in which this answer buttresses the claim that determinism is fundamentally irrelevant for responsibility. Section VIII briefly summarises the chapter and looks ahead to the next and final chapter.

II. The Explanatory Task

For philosophical understanding, it is *necessary* that we add to our enquiry an answer to the unanswered question; we should consider Strawson’s account incomplete without it.¹⁸⁰ In any case, that is Strawson’s own opinion: “full understanding of a concept is not achieved until this enquiry is added”. According to him, unless we “explain [...] why it is that we have such concepts and type of discourse as we do” (2011a, p. 86) our philosophical understanding is incomplete.¹⁸¹ If we are to “recover from the facts as we know them as sense of what we mean, i.e., of *all* we mean, when” (2008a, p. 24) using the concept

¹⁸⁰ This is on the assumption that the answer to be given is, as I’ll argue, provided by the explanatory task of the philosopher. Doubts may be had about this point. Is asking for the ‘general causes’ to ask for a *philosophical explanation* (of some specific kind), rather than to ask for a *causal story* of why we have reactive attitudes? It is not impossible that the latter is what Strawson actually had in mind. This of course does not mean that the answer to be developed here (which is an exercise of ‘the explanatory task’) is not informative or particularly informative of Strawson’s project, as I will argue it is. But it is not inconceivable that Strawson had something else in mind when posing this question, and others like it elsewhere. However, to place this question centre stage, even if it is neglected, would be odd if it is not of any particular concern. And Strawson did not pay the causal story any particular concern (not, of course, because it is of *no* relevance). At the very least, the way Strawson *should* have answered the question is, broadly, as it is answered here, and not by a causal story.

¹⁸¹ Strawson’s opinion is, perhaps, in characteristic fashion, somewhat less determinate than this presentation may make it seem. He thinks that “[i]t may reasonably be maintained, or ruled, that” the explanatory task is necessary for fully understanding a concept (2011a, p. 86); that is, it “*may* be required of the philosopher” (2011b, p. 36, emphasis added). Still, at the very least, he does think that “a fuller understanding [...] *may* be gained by enquiring into [types of discourse and concept] foundation in natural facts” (2011a, p. 90, emphasis added).

of responsibility, we need to attend to the facts as we know them also from this direction.

we should consider seriously the origin—the fount and origin—of the distinction we draw in practice between cases where responsibility is ascribed and cases where it is not. (Strawson, 1998b, p. 260)

What we need to do is what Strawson calls *the explanatory task* of the philosopher; until this has been fulfilled, full understanding of the concept cannot have been achieved.

For the Minimal Optimist Reading proposed in this thesis—focusing on that “core of human thinking which has no history” (1959, p. 10)—the explanatory task is particularly important, since the descriptive metaphysician’s task is itself “interlaced with” (*inter alia*) the explanatory task, and inextricably so—it “cannot be detached from” it (Strawson, 2011a, p. 89). To not seek an answer and see how this answer affects our understanding of the concept, our sense of what we mean when using it, should start to seem, both philosophically (if we think Strawson has a point here) and (regardless of our sympathies) interpretatively, rather neglectful. The question *is* there, in plain sight. How should we answer it?

What Strawson calls ‘the explanatory task’ of the philosopher is to ask “why it is that we have such concepts and types of discourse as we do” and then to “show the natural foundations of our logical, conceptual apparatus, in the way things happen in the world, and in our own natures” (Strawson, 2011a, p. 86; also, 2011b, p. 36).¹⁸² Differently put, the task is to “understand the foundation

¹⁸² Strawson introduces the distinction between five different philosophical tasks in some early methodological papers: ‘Construction and Analysis’ (2011b), ‘Analysis, Science, and Metaphysics’ (2011a) and ‘Carnap’s View on Constructed Systems versus Natural Languages in Analytic Philosophy’ (1963). The latter two papers are largely overlapping, especially on the points relevant here. ‘Analysis, Science, and Metaphysics’ is a translation by Richard Rorty (1992), from French, of Strawson’s contribution to the Royaumont Colloquium, and for overlapping passages Rorty has used the same wording in his translation as that which appears in ‘Carnap’s View ...’. It is of historical and, for the present reading, interpretative interest to note that, though it is a bit unclear, it seems like (see, Overgaard, 2010) the French original of ‘Analysis, Science and Metaphysics’ was presented in 1958, not long before Strawson wrote ‘Freedom and Resentment’ and ‘Social Morality and Individual Ideal’ (ca. 1960). ‘Carnap’s View ...’, though published first 1963, was “[c]ompleted to a deadline in 1954” (Strawson, 1998a, p. 10). Three of Strawson’s five tasks are *descriptive* and the other two *imaginative*. We are primarily concerned with the imaginative here: the *explanatory* task and the *creative* task. One descriptive task is *therapeutic*, aiming to resolve philosophical paradoxes by exposing “the full logical workings of the distorted concepts” (Strawson, 2011b, p. 35). The second descriptive task is to do the same, but *for its own sake*, and hence perhaps more systematically than

of our concepts in natural facts” (2011a, p. 89). What we are seeking to uncover is, first and foremost, “how our concepts, as they are, are rooted in the world, as it is” (2011b, p. 36). Strawson gives an example of the schematic form that such an explanatory enquiry may take:

if things (or we) were different in such-and-such ways, then we might lack such-and-such concepts or types of discourse; or have such-and-such others; or might accord a subordinate place to some which are now central, and a central place to others; or the concept we have might be different in such-and-such ways (Strawson, 2011a, p. 86)

Here, Strawson echoes Wittgenstein’s remarks on “the correspondence between concepts and very general facts of nature” (2009b, §365).¹⁸³ Just as Wittgenstein, Strawson notes that the kind of explanation we are seeking “is not an historical enquiry” (2011a, p. 86). Supposedly, taking up Wittgenstein’s note, this is because “we can also invent fictitious natural history for our purposes” (Wittgenstein, 2009b, §365). Fictionalisation need not imply, and does not for Wittgenstein, a denial of the fact that “concept formation can be explained by facts of nature” (Wittgenstein, 2009b, §365). But it does, however, raise a question about what ‘our purposes’ are in doing philosophy; and, of course, about what Strawson’s philosophical purposes are.¹⁸⁴

Strawson cursorily divides the relevant natural facts into two categories: facts pertaining to “the nature of the world” and facts pertaining to “our own nature” (2011b, p. 36). We should not make too much of this contrast: recall, “The set of ideas, or schemes of thought, employed by human beings reflect, of course, *their nature*”—and this means, as Strawson elaborates, that the set

(philosophical) therapy requires. The third descriptive task is the *descriptive metaphysician’s task*, of exposing the most general and fundamental concepts.

¹⁸³ “If anyone believes that certain concept are absolutely the correct ones, and that having different ones would mean not realizing something that we realize — then 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 our usual ones will become intelligible to him” (Wittgenstein, 2009b, §366). We find similar remarks by Strawson also elsewhere; for example, in an already considered passage: “At the stage of conceptual self-consciousness which is philosophical reflection, people may, among other things, conceive of variation in the character of their own situations and needs and discuss intelligibly the ways in which the schemes of thought might be adapted to such variations” (Strawson, 1966, p. 65f.). For the affinity between Wittgenstein and Strawson on this point, see Bengtson (2019) and Queloz (2021a, p. 29).

¹⁸⁴ On Strawson’s conception of philosophy and general methodology, see Heyndels (2019, 2020, 2022b); also Glock (2012, 2017, 2022) and De Mesel (2022b).

reflects “their needs *and* their situation” (1966, p. 65, emphasis added). Nor should we think that the contrast may always be neatly posed—it cannot. Though it is admittedly a vague distinction, it is not therefore unworkable or unimportant (Strawson, 2011a).

One reason why the distinction is important has to do with Strawson’s interest in the most general and fundamental features of our conceptual scheme. In light of the diversity of human nature—that our conceptual schemes “are not static schemes” but change with (*inter alia* but saliently) “the development of social forms” (Strawson, 1966, p. 65)—we cannot rest content with explaining why the conceptual scheme we have is as it is. The philosopher’s imagination needs to engage with natural facts not only in execution of the explanatory task, but also through *the creative task*. That is, we need to consider, also, “how, without the nature of the world being fundamentally different, we might nevertheless view it through the medium of a different conceptual apparatus, might conduct our discourse about it in forms different from, though related to, those which we actually use” (Strawson, 2011b, p. 36). It is for this task, the creative task, that the distinction between ‘our own nature’ and ‘the nature of the world’ is relevant. This further imaginative task aids the descriptive task of mapping, not just our *actual* conceptual scheme, but the *fundamental* and *most general* framework of any human conceptual scheme.¹⁸⁵

While the process of unearthing the natural foundations of our concepts may render explicit the *non-arbitrariness* of our conceptual scheme, presenting it as an intelligible upshot of how we and the rest of the world are like (Heyndels, 2019),¹⁸⁶ the creative task may award an appreciation of the *contingency* of our conceptual scheme. When the two are added together, *certain* features come into focus—viz., the most general and basic concepts, those “which, in their fundamental character, change not at all” (Strawson, 1959, p. 10). In understanding *why* we have the conceptual scheme we have, we also see how the set of ideas that we possibly could have is *limited*. We are thus brought to

¹⁸⁵ As Heyndels notes, the *positive* function of imagining conceptual variation in relation to the same world is primarily served when this task is “secondary to the philosopher’s descriptive task” (2019, p. 411). Still, Strawson’s uninterest in ‘*revisionary* metaphysics’ should not be exaggerated (cf., De Mesel, 2022b).

¹⁸⁶ This suggestion finds support, for example, in considering Salomon’s (1957) suggestion that Strawson (1952) presents inductive reasoning as *mere convention* or as a matter of an *arbitrary choice*; to which, as we’ve seen, Strawson’s (1958) response is that there is no arbitrary choice where there is Nature. See also his “Reply to Ernest Sosa” (1998c): “if we are moved to ask ‘How does all this come about?’, we cannot improve on Hume’s answer: ‘Nature’” (p. 371).

appreciate the inescapability or necessity of certain features of our conceptual scheme.

Focusing on the Inescapability Claim—on the Minimal Optimist construal: that the *concept* of responsibility is humanly necessary, but no particular *conception* of responsibility is—the value of the explanatory task for Strawson’s approach in ‘Freedom and Resentment’ is evident. The philosopher’s imagination—involving both the explanatory task and the creative task—is crucial for the descriptive task of exposing precisely those non-historical concepts among which, the claim is, we are to count ‘responsibility’.

III. Justification Contra Explanation

There is a further reason for taking up the explanatory task that is not evidently present if we only consider ‘Freedom and Resentment’. The secondary literature on Strawson’s way with responsibility scepticism is rife with dissatisfaction about his insistence that, at a certain level, no further justification is to be given; that what we end up encountering are just natural facts about us. That, “[t]he existence of the general framework of attitudes itself is something we are given with the fact of human society. As a whole, it neither calls for, nor permits, an external ‘rational’ justification” (Strawson, 2008a, p. 25).

In response to Strawson’s shunning of the invitation to supply an external justification, it is said that his approach is “inadequate” because the “worry about the metaphysical grounding of our responsibility practices” persists unanswered (McGeer, 2014, p. 65),¹⁸⁷ that the approach is “relentlessly naturalistic” (Coates, 2017, p. 804), or, most pointedly, that it suffers from a kind of “naturalistic complacency” (McGeer, 2014, p. 88).¹⁸⁸ For the rehabilitative purpose of this interpretation, we should take these accusations seriously. This does not mean that we should endorse the sentiments underlying these accusations. What commentators tend to do—driven as they are by ‘the Archimedean urge’ (Srinivasan, 2015)—is to try to reach further

¹⁸⁷ See also Pears’s remarks to the effect that “We may well wonder how much weight this kind of naturalism can be expected to bear” (1998, p. 248).

¹⁸⁸ McGeer adds that “Strawson’s cavalier response not only appears off topic, it further smacks of a kind of moral complacency” (2014, p. 89). A similar criticism is, as summed up by McKenna and Russell (2008a, p. 10), that Strawson’s account is “conservative”.

and conjure up some *justificatory* story of our responsibility practices. In pursuing an enterprise antithetical to Strawson's (2008c, p. 19) outlook, they (supposedly), as we may put it (cf. Strawson, 2008c, p. 19), find it too "difficult to begin at the beginning" and thus "try to go further back" (Wittgenstein, 1969, §471). The absence of a rational foundation for our practices taken *in toto* may spur many different responses, and a common one, a perfectly understandable one, but a wholly unacceptable one for our purposes, is to defy the anti-foundationalist idea and seek an external justification.

However, to paraphrase, might we not induce the critic to give up saying this by giving the Minimal Optimist something more to say? The key to our answer is in fact already in Strawson's arsenal. In response to manifestations of this Archimedean urge, Strawson makes use of a distinction between 'justification' and 'explanation'. For example,

To say that this [... fact of human nature ...] does not need to be *argued for* [i.e., *justified*] is not to say that it does not need to be *explained*. We may discuss its natural sources (Strawson, 1961, p. 12, emphasis added)

As we've seen, the key insight, that "sceptical doubts are not to be met by argument", Strawson derives both from Wittgenstein and from Hume. And Hume (as Strawson (2008c, p. 9, e.g.,) notes) also hands us the relevant distinction precisely when, with the other hand, he dismisses "the reasoned rebuttal which the sceptic pervasively invites" (Strawson, 2008c, p. 16):

We may well ask, *What causes induce us to believe in the existence of body?* But 'tis in vain to ask, *Whether there be body or not?* That is a point, which we must take for granted in all our reasoning. (*Treatise* I.IV.II p. 125)¹⁸⁹

This use of the distinction between, on the one hand, (rational) *justification* and, on the other, (naturalistic) *explanation* not only shows that the distinction is recognised by Strawson—which is important for the interpretative basis of the present reconstruction. This use moreover indicates the distinction's philosophical purpose: explanation may help (in some underspecified way) *precisely* when justification cannot be, and is not to be, sought. Recognising the limits of Reason, we are not debarred from asking (quite the contrary—Strawson encourages us to ask) *why* it is that we live by this concept—*why* do

¹⁸⁹ The same goes for our proneness to reactive attitudes as for our belief in the existence of body (Strawson, 2008c, p. 26) and for induction (Strawson, 2008a, n. 7).

we have it?¹⁹⁰ The end of our legitimate justificatory pursuit is not coincidental with the end of our philosophical pursuit for greater conceptual self-understanding.¹⁹¹

Strawson's use of explanation in terms of "natural sources" rather than justification, its use precisely where justification comes to a halt, and the explanation's potential to, in light of the widespread sense of dissatisfaction with Strawson's insistence on the limits of justification, in some sense alleviate the dissatisfaction—all of these factors, taken together, add stress to the value of the explanatory task.

IV. An Answer to the Unanswered Question

As so happens, there is a kind of naturalistic explanation that makes for a promising answer to the unanswered question of 'Freedom and Resentment'. The kind of explanation I propose we supplement Strawson's approach with is that of a *pragmatic genealogy*.¹⁹²

The historical developments of this method, as Matthieu Queloz (2021a) has shown, may be traced from David Hume (2011) and Friedrich Nietzsche (1986) to Edward Craig (1990), Bernard Williams (2002), and Miranda Fricker (2007). The emerging argument of this chapter is that we should read Strawson's 'Freedom and Resentment' *as if* he belonged to this group of genealogists, even if he never in fact developed any pragmatic genealogy.¹⁹³

¹⁹⁰ Indeed, that this question is given a crucial role in Strawson's meta-philosophy at least complicates the picture of him as complacent or conservative.

¹⁹¹ Conceptual self-understanding or self-consciousness is, according to Strawson, the aim of philosophy (Pivčević, 1989, p. 4; Strawson, 2011e, p. 225; 2019, p. 34). Heyndels (2020, esp. p. 3) emphasises this, taking it to be Strawson's *central* meta-philosophical idea.

¹⁹² In the following, I draw on Queloz's explication of the method of pragmatic genealogy, as well as Williams' use of it. The purpose of this chapter is not to assess or develop the methodology, but to illustrate the fruitfulness of reading Strawson *as if* this is the answer he would have provided if he had answered the unanswered question. More precisely, the purpose is (i) to show the congeniality of the method of pragmatic genealogy to Strawson's method, his philosophical aims and concerns and (ii) to suggest that, whatever Strawson would have answered, we (especially we who are intrigued by Strawson's general approach, as it has been presented here) have reason to use the method of pragmatic genealogy in response to the concerns raised by some of Strawson's critics.

¹⁹³ Though, as we will see below, not too far from it (cf., Strawson, 2008d). As already noted, Benjamin De Mesel also explores a kind of genealogical explanation as implicit in Strawson's approach, even if it is not quite of the same kind as the one developed here.

However, while not a pragmatic genealogist, the central idea of this approach—that general facts about human beings render practically necessary certain concepts—is present in Strawson’s work. The thought is surely not unique to him. Several philosophers in Strawson’s intellectual milieu also held that certain concepts are necessary in any human society in virtue of facts about the needs and situations basic to the human condition.¹⁹⁴

Isaiah Berlin, for example, thought that there is some minimum structure of moral values “without which human societies would disintegrate and from which [...] men cannot depart without perishing” (2017, p. 234). Stuart Hampshire also thought that some virtues are essential to any human form of life—even though the *form* these virtues take, their relation to other virtues and much else, undoubtedly differ greatly between societies. The structure underlying these essential virtues “has no supernatural source, it must be recognized by rational inquiry as having its origin in nature and, specifically, in human nature: that is, in constant human needs and interests” (Hampshire, 1983, p. 128). Similarly, Herbert Hart (2012), one of Strawson’s close friends,¹⁹⁵ argued that certain rules of conduct are necessary for any social organization given the practical needs human beings invariably have.

The common idea of these philosophers is that basic human needs and situations render practically necessary certain concepts for any human society. Next, we’ll see how this idea may be illuminatingly extended by the method of pragmatic genealogy.¹⁹⁶

However, in a forthcoming paper, De Mesel (forthcoming) explicitly develops a *pragmatic* genealogy in relation to Strawson’s approach which is along the same lines as the one presented here.

¹⁹⁴ I thank Matthieu Queloz for directing me to this common thread of thought (cf., Queloz, 2023, p. 7). For discussion of this idea (especially as it relates to political philosophy) in Berlin, Hampshire, and Williams, see Hall (2020).

¹⁹⁵ Strawson “liked him [Hart] tremendously” (Lacey, 2006, p. 262). Their friendship is recognisable by the fact that Hart is one of few given acknowledgement in the prefaces of Strawson’s books, and the only one thanked both in the preface of *Individuals* (1959, p. 3) and the preface of *The Bounds of Sense* (2019, p. xix).

¹⁹⁶ As it explicitly is in Williams’ (esp., 2002) work.

The Method of Pragmatic Genealogy

A pragmatic genealogy is a dynamic philosophical model of the practical origins of a conceptual practice.¹⁹⁷ The object is to understand the *function* of a particular practice (the *target-practice*) by asking why concept-users like us would go in for that practice in the first place. The model of a pragmatic genealogy illuminates a certain kind of dependence structure of the target-practice, namely the *instrumental* dependence structure: “the ways in which conceptual practices are instrumental to the satisfaction of concept-users’ needs” (Queloz, 2021a, p. 50).¹⁹⁸ We should ask ‘What is the point of the practice?’ and then, based on the functional hypothesis we arrive at, construct a model in which a prototype of the target-practice (the *proto-practice*) is shown to solve a basic problem that people like us would face.

We start from a State-of-Nature condition where human beings, in virtue simply of some maximally generic needs and situations, would face a basic problem that the proto-practice presents itself as a salient solution to.

This State-of-Nature condition is, importantly, not to be mistaken for a description of an *actual* (pre-historical) period of hominin life. What we are here providing, in line with Strawson’s vision, “is not an historical enquiry” (1963, p. 516).¹⁹⁹ The genealogy initially models a highly *idealised* condition, the State-of-Nature, and then *de-idealises* this model to bring us closer to the target-practice. The pragmatic genealogy thus combines both *fiction* and *history*.

Fictionalisation here means simply that we make use of idealisation in constructing our State-of-Nature condition, which implies a certain divergence from reality, as any idealisation does.²⁰⁰ The genealogy then de-idealises in two

¹⁹⁷ Strawson’s association with ordinary language philosophy should not lead us to think that model-building is alien to Strawson’s methodology; see De Mesel (2022b) for a perceptive discussion of this issue. As Queloz (2021a, p. 29, n. 16) notes, Strawson seems to allow for model-building *specifically* in relation to exercises in philosophical imagination: “to understand the foundation of our concepts in natural facts, and to envisage alternative possibilities, it is not enough to have a sharp eye for linguistic actualities” (Strawson, 1963, p. 517).

¹⁹⁸ Typically, a genealogy is a developmental explanation of *cultural* phenomena (see, Queloz, 2021a, p. 52, n. 5).

¹⁹⁹ Recall, “we can also invent fictitious natural history for our purposes” (Wittgenstein, 2009b, §365).

²⁰⁰ Different forms of fictionalisation include ‘Aristotelian’ idealisation “that abstract away from the particular without distorting it” and ‘Galilean’ idealisation “that distorts reality in order

phases. First, as further generic needs are anticipatable already *within* the State-of-Nature, we see how the proto-practice needs to be elaborated already in this schematic condition to a *generic version* of the target-practice. Second, by incrementally factoring in socio-historically local needs, we see how the generic version of the target-practice must be elaborated further as to answer also to these particular needs, bringing us closer to the actual conceptual practice that we, here and now, have. The closer this process of de-idealisation brings the model's practice to our actual practice, the stronger is the genealogist's claim to having identified the practical origins of the practice.²⁰¹ *History*, if only at a rather abstract level, is necessary for the final de-idealization that gets us beyond the State-of-Nature and closer to the target-practice.

If the genealogy reveals that a conceptual practice solves a very basic problem—a problem that *any* group of humans would face, since the problem presents itself already if we assume only that these humans have some very generic needs in some very common situations—then the practice is revealed to be *practically necessary* for beings like us (human beings). Given that these needs and these circumstances indeed are part of our genealogical explanation already at the highly idealised stage of the State-of-Nature condition—this fiction of the philosophical imagination—the concepts that are practically necessary already at this stage may be considered *non-historical*. The presence of such a concept in any *actual* human society is explicable in terms of a practical necessity that is *antecedent to the actuality* (to the 'in-history', if you like) of any actual human society. The pragmatic genealogy, with its fictional State-of-Nature condition, makes for a congenial way of elucidating that elusive idea that we find in Strawson—of certain concepts being, in their fundamental form, without history.

to illuminate it" (Queloz, 2021a, p. 13), as well as the combination of these: 'caricature' (Queloz, 2021a, p. 50).

²⁰¹ The functional hypothesis, while a hypothesis, is not an arbitrary matter. It is to be *retroactively vindicated* by the explanatory power it demonstrated through the ensuing genealogy (Queloz, 2021a, p. 51).

A Pragmatic Genealogy of Blame

An example will help to make this clearer. Take *blame*.²⁰² What is the point of blame?

Instrumentalist and functionalist accounts of blame have provided a variety of answers to this question.²⁰³ Blame is instrumental, generally speaking, for “regulating behaviour in socially desirable ways” (Strawson, 2008a, p. 2). But this is far too unspecific to characterise ‘blame’ as opposed to other means of regulating behaviour. More specifically, the way in which blame is instrumental for regulating behaviour is by way of being a reaction (typically, an emotional reaction) to the behaviour of other agents which manifests an insufficient responsiveness to some (moral) reason, thereby rendering the blamee more responsive to the relevant reason and aligning their ethical sensibilities with the prevailing or socially desirable ethical sensibilities.²⁰⁴

Let’s take this to be our *functional hypothesis* about our practice of blame.

From this, we may construct a prototype of our practice that is *purely instrumental*. According to our proto-practice, the *only* reason to blame someone is the instrumental reason that blame shapes their ethical sensibilities in desirable ways by making them more responsive to some reason that their behaviour shows an insufficient responsiveness to. What basic problem would

²⁰² Besides the anchoring in Strawson’s work, the section largely draws on Queloz’s (2021b) account. See also Queloz (2018, 2020, 2021a), Williams (1972, pp. 123, 139f.; 1985, 1993, 1995a, 1995b, 2002), and Sliwa (2019).

²⁰³ Instrumentalism about blame (or responsibility more generally) is the view that the practice or the shape of the practice or any token of the practice is to be justified in terms of the value of its outcomes. Some noteworthy instrumentalist accounts include McGeer (2012, 2013, 2019), Milam (2021), Jefferson (2019), and Vargas (2013); for overview, see Vargas (2022). Here, ‘functionalism’ is used in contrast to ‘instrumentalism’. A view is functionalist if it thinks that the function of blame is philosophically relevant for an account of blame but does not subscribe to the instrumentalist claim that the function is what blame should be justified in terms of. Some noteworthy functionalist accounts of this kind are McKenna (2012) and Macnamara (2015a, 2015b). ‘Functionalist’, as I’m using the term, is not necessarily what Shoemaker (2020, p. 221) calls ‘a function approach’, which he contrasts to a content approach; for example, McKenna’s (2012) conversationalist theory counts as a functionalist one on the present proposal, but a content approach on Shoemaker’s presentation.

²⁰⁴ Strawson agrees with Prasad (1995) that “when we give voice or effect to our moral reactions [...] our purpose is normally not merely expressive [...] or merely retributive [...] Our purpose is, at least partly, to call the reprov’d agent’s own attention to the moral quality of his act, so that he, seeing himself as a moral agent, may be moved *freely* to amend his ways” (Strawson, 1995, p. 430f., original emphasis) See also Williams (1985, pp. 213-216; 1995a, p. 16; 1995b, p. 40f.) and Fricker (2016).

this proto-practice be a salient solution to? What generic needs and situations combine to form that problem?

Strawson's claim is that "[t]he existence of the general framework of attitudes itself is something we are given with the fact of human society" (Strawson, 2008a, p. 25). Why is that? Let's imagine, while idealising away as much as possible from our own particular conditions, a small society of human beings, with a common language, but without a practice of blame. Reformulating our question slightly, we may now ask: Why would already this society of human beings come to have a practice of blame?

Drawing on Strawson's 'Social Morality and Individual Ideal' (2008d), we get a sense of why. According to Strawson, "it is a condition of the existence of any form of social organization, of any human community, that certain expectations of behaviour on the part of its members should be pretty regularly fulfilled" (2008d, p. 33f.). There are both individual and social needs that explain why that is. For the individuals, each has some need for stability beyond that secured by the spontaneous altruism and sympathy of her immediate family and friends. For example, she has, for pursuing her own ends, a need to be able to rely on others' behaviours and, for more advanced pursuits, a need to cooperate with them in pursuing those ends.

For the society, for its perdurance, there need be something holding it together; there need be some minimal ethical sensibility generally shared by its members if the society itself is not to unravel under destabilising conflict.²⁰⁵ As individual members' interests conflict, some basic ethical sensibility needs to be shared between them if such conflict is not to undermine the "system of human relationships" itself. This amounts to a "minimal interpretation of morality" that understands morality in the first instance as "a kind of public convenience": a practical necessity, given the needs of individuals and society (Strawson, 2008d, p. 33f.). Already this small society of human beings, without us assuming anything very specific about it or its members, would have a practical need for some minimal form of morality, for some generally shared ethical sensibility.²⁰⁶

If some ethical sensibility needs to be generally shared within the society, *unaligned* individuals—new individuals (most notably children) and individuals that deviate from the ethical practice in problematic ways—present

²⁰⁵ How much in the way of a shared ethical sensibility is actually necessary depends on several factors. Saliently, for us that is, it depends on the political arrangements of the society.

²⁰⁶ As Strawson sums it up at one point: "for who could exist at all, or pursue any aim, except in some form of society? And there is no form of society without rules, without some system of socially sanctioned demands on its members" (2008d, p. 38).

the society with a practical problem. Enter blame: it is as a technique for alignment and realignment that the proto-practice of blame presents itself as a salient solution to this practical problem.²⁰⁷ The problem of unaligned individuals presents the human beings in our State-of-Nature with an instrumental reason to make these individuals more responsive to reasons that their behaviour shows an insufficient responsiveness to. This basic problem, to which the purely instrumental version of blame forms a solution, follows already on the assumption of some very generic needs of individual human beings and human society. This is the first step of our pragmatic genealogy of the practice of blame.

If we stick with the blamer's perspective, that of the individual trying to (re-)align unaligned individuals, it is not obvious why such a purely instrumental practice of blame would not be able to fulfil the relevant function. Importantly, since blame regulates behaviour *by* shaping agents' ethical sensibilities, it will be effective only if it in fact gets agents to really take the relevant considerations to be reasons. Blame is effective, that is, if it fosters dispositions in agents to feel, e.g., remorse, guilt, and shame when they fail to be properly responsive to the relevant considerations. This, however, does not immediately imply that the blamer cannot employ a purely instrumental understanding of blame's rationale: the blamer may try to instil a sensitivity to certain reasons in the blamee without assuming that the failure to attend to these reasons *itself* makes blame appropriate. The blamer may be *hypocritical*, pretending that the considerations which the blamee is said to have failed to be properly responsive to in fact justifies blame if the agent is not properly responsive to them.²⁰⁸ Though the blamer herself does not actually think that some consideration is itself such that failures to be responsive to it justifies blame, the blamer thinks that if people *take* the consideration to be a reason that warrants blame when neglected, then this will be for the best, and the blamer therefore communicates the sense that the consideration is important by pretending to blame because it has been neglected.

If purely instrumental blame is generally to be effective, the blamer needs to be hypocritical, as blame which is seen as unjustified will typically be met

²⁰⁷ There are other solutions to this problem; some of them get at (re-)alignment by other means (such as propaganda) and some of them get at stability despite misalignment (such as punishment or imprisonment). Importantly, when blame solves this problem, it is not only blame that is necessary: it is only as part of a concrete ethical life that blame can serve its function (B. Williams, 1995a, p. 16; 1995b, p. 41).

²⁰⁸ 'Properly' here allows for the fact that the blamee might have treated the fact as a reason while not treating it in the way that, given the reasons that it is, it should be treated (e.g., to use a common metaphor, by giving insufficient weight to it).

with resentment rather than recognition. Notably, it is not enough if the blame only gets the blamee to abstain from the behaviour in the presence of the people that blame them; that will not make them more responsive to the relevant considerations, will not (re-)align their ethical sensibility. Blame will not have fulfilled its function if that is all that it achieves.

Since any token of blame will serve the function of the proto-practice only if it succeeds in inculcating a concern for the relevant consideration in the blamee, the blamee cannot remain unresponsive and simultaneously consider the blame justified. The blamee “cannot coherently think: ‘Though I don’t care at all about *X* myself, things go better if people in general are responsive to *X*, so since, by blaming me, *A* has rendered me more responsive to *X*, *A* was instrumentally justified in blaming me” (Queloz, 2021b, p. 1365). The blamer must really get the blamee to become responsive to the relevant consideration, get them to care about *X*, by transmitting a sense that *X* really merits concern. This sense—that *X should* be considered—is necessary if the agent is to experience self-directed reactive attitudes such as remorse, guilt, or shame when they are not responsive to the consideration. To have these self-reactive attitudes is expressive of having an ethical sensibility; that, in a particular form, is thus a crucial part of the instrumental value of blame. The blamee being prone to these reactions is a success condition for the functionality of blame.

Furthermore—now introducing a dynamic aspect to our very schematic society—it is plausible to assume that in any society, agents will tend to *alternate roles* within the practice, sometimes being blamers and sometimes being blamees. Even if hypocrisy and pretence may work given the assumption that those blaming and those being blamed are not the same people,²⁰⁹ the functionality of the practice will be undermined by the fact that people alternate roles. This fact makes the practice unstable. The agent that is blamed will know that blame is only justified if it regulates behaviour in socially desirable ways by getting them to care about the relevant considerations. Hypocrisy won’t save the practice on the assumption that parties will alternate roles, since this will eventually be revealed as just that: hypocrisy. Since blame that is considered unjustified is typically not effective, the proto-practice will not fulfil its function. In the proto-practice, there is nothing else to appeal to but the instrumental value of having people in general caring about *X*, but this does not justify any instance of blame unless it actually gets the blamee to care. In order for the proto-practice to serve the function of blame, it must therefore develop beyond its purely instrumental form; it must develop such that *other reasons* than the efficacy of the blame may justify blame. This feature of the

²⁰⁹ See Williams (1985; 1995a, p. 15; 1995c).

practice, the non-instrumental character of blame, turns out to be a *precondition of functionality* for the practice of blame (Queloz, 2021b, p. 1365).

Already within our State-of-Nature condition, we see why our practice of blame has the fundamental structure that it has. On our pragmatic genealogy, blame is understood as a *self-effacingly functional* practice. It is functional, but only if the practice is sustained by motives and reasons that are autonomous, i.e., not dependent on the functionality of the practice in any particular instance of blame, and, lastly, the explanation for why the practice is autonomous is that its functionality depends on its being autonomous.²¹⁰ A generic version of the target-practice (of our practice of blame) is anticipatable already from within a dynamic model of a human society, assuming only that it and its members have some very generic needs. On these assumptions, we see that *any* human society would face the same basic problem and come to develop *some* practice of blame—that is, some socio-historical development of the generic version of the practice of blame. For our present purposes, this initial de-idealisation, within the State-of-Nature condition, suffices.

The Concept of Responsibility

The example of blame bears on what the concept of responsibility is, since one feature of responsibility is that of an action giving rise to a response, and to engage in blame is one such response. It is important that it is only one such response. If we are to vindicate Minimal Optimism, it is nevertheless a very pertinent one. It is typically this element of our responsibility practices that encourages pessimism and genuine scepticism about responsibility.

Having illustrated the method of pragmatic genealogy with the example of blame, we may supplement our understanding of the practical origins of the concept of responsibility with a somewhat more *ad hoc* genealogy of some other central features of that concept. For one thing, people everywhere and anywhere (i.e., already in the State-of-Nature) need to have the concept of *intention*; they need to be able to distinguish between actions done intentionally and actions done unintentionally. Williams (1993, p. 50ff.) exemplifies this with the case of Odysseus and Telemachus being surrounded by Penelope's suitors. Even though they had prudently stored away the suitors'

²¹⁰ That is, the practice exhibits what Queloz (2018, p. 14; 2021a, p. 57; 2021b, p. 1366) calls the FADE structure: *functional*, *autonomous*, the functionality is *dependent* on the autonomy, and there is an *explanatory connection* between its autonomy and the functional dependence on autonomy.

armour before the assault, Odysseus sees with chock how the suitors are now handing out spears. Odysseus wonders how this could be, and Telemachus reveals that it was he who left the door ajar; “One of them was a better observer than I”, he tells Odysseus. What he conveys to Odysseus is not only that he left the door open. What he conveys is that he did so *unintentionally*. This is enormously important for Odysseus: if Telemachus had *intentionally* left the door open, Odysseus’s current situation would probably be much worse than it in fact is. If nothing else but to know what to expect from other people, people living together must take an interest in the intentions of other people. It is practically necessary for human beings everywhere and anywhere to make some distinction between intentional and unintentional actions.

Strawson gives a similar example when emphasising “how much we actually mind, how much it matters to us, whether the actions of other people [...] reflect attitudes towards us of goodwill, affection, or esteem on the one hand or contempt, indifference, or malevolence on the other” (Strawson, 2008a, p. 6f.)

If someone treads on my hand accidentally, while trying to help me, the pain may be no less acute than if he treads on it in contemptuous disregard of my existence or with a malevolent wish to injure me. But I shall generally feel in the second case a kind and degree of resentment that I shall not feel in the first. (Strawson, 2008a, p. 7)

This exemplifies what Watson (2008, p. 119) dubbed ‘the basic concern’. What a pragmatic genealogy emphasises, beyond this, is the needs that themselves explain why it actually matters to us, why we mind so much, what other peoples’ attitude towards us is—i.e., why we have the basic concern.²¹¹

Strawson also holds that the reactive attitudes are sensitive to whether the agent’s *state of mind* when acting was normal. Typically, the abnormality of the agent is itself treated by commentators as incidental or irrelevant; what Strawson must have meant is that the agent is in some sense *incapacitated*. We might say that the agent is incapacitated for engaging in *normal* human relationships—that the agent is abnormal in that sense.²¹² However, there is

²¹¹ Watson is getting at this too: “These stances and responses *are expressions of certain rudimentary needs and aversions*” (2008, p. 117, emphasis added). I thank Maria Alvarez for pressing me to get clearer on how ‘the basic concern’ figures in the present explanation.

²¹² See, for example, McGeer (2014, p. 75): “What makes [someone who is cognitively and affectively abnormal] an ‘inappropriate object’ of this normative demand? Strawson’s thought is clear: their cognitive/affective handicap either makes them *incapable* of understanding the kind of demand expressed in our reactive attitudes or it makes them

another way—in the present context, a more straightforward way—to understand Strawson’s point.

An explanation in terms of the needs that render it practically rational for beings like us to want the concept if they did not already have it allows us to see why the reactive attitudes would be mollified or modified when an action is done in an abnormal state of mind. For, everywhere and anywhere, people have an interest in knowing how the actions of other people, whether intentional or unintentional, fit in with the agent’s other interests and plans. Whether the state of mind of the agent is normal or abnormal, as Queloz puts it, “helps us separate the exceptional from the expectable” (2022b, p. 190). For anyone who is to live with the person, to be able to draw such a distinction is of great interest. Strawson emphasises this point:

We normally have to deal with him under normal stresses; so we shall not feel towards him, when he acts as he does under abnormal stresses, as we should have felt towards him had he acted as he did under normal stresses. (Strawson, 2008a, p. 9)

To be able to draw these distinctions—between intentional and unintentional actions, and between actions done in a normal and an abnormal state of mind—is practically necessary if human beings are to be able to live together. For any human society, there is a practical need for a concept of responsibility that allows its members to differentiate between actions in something like these terms.²¹³

What this way of answering Strawson’s unanswered question suggests so far is that the concept of responsibility involves not only the idea of a response, but also the idea of intention and of state of mind.²¹⁴ These ideas take different forms and are variously related to each other on different conceptions of responsibility. But since they follow, as Williams puts it, “simply from some

incapable of living up to that demand. Hence, they are unfit to be treated as ‘participants’ in our shared moral practices”.

²¹³ This is a bit too quick. That we have a need for this distinction does not by itself mean that we have a need for a concept of responsibility that is sensitive to this distinction, for why must that need be served by one concept; in particular, the concept of responsibility? A basic point, a reminder really, is that the concept of responsibility *is* sensitive to this distinction. This is so even if we haven’t fully unearthed the practical origins of that concept. For one explanation of why the needs here discussed are merged under the same concept, see Queloz (2022a, pp. 1602-1608).

²¹⁴ Williams (1993, p. 55) adds ‘cause’ to ‘intention’, ‘state’, and ‘response’.

universal banalities”, it is very unlikely that any human society could exist without such a concept.

V. The Answer Strawson Should Have Given

The method of pragmatic genealogy provides an explanation in terms of “the natural foundations of our logical, conceptual apparatus” that “is not an historical enquiry” (Strawson, 1963, p. 516) but essentially a product of “the philosophical imagination” (Strawson, 2011b, p. 36). This method, as exemplified with the case of blame and (in less explicit form) with respect to further elements of the concept of responsibility, helpfully illuminates what answer Strawson should have given to the unanswered question of ‘Freedom and Resentment’, for the following reasons:

The explanation substantiates a crucial sense in which the practice of blame and the concept of responsibility are inescapable: they are *practically necessary* for beings like us. The Inescapability Claim is not that to have the concept is psychological necessary. Neither is the claim only that they are conceptual necessity. That we would do without the concept or responsibility is, “for us as we are, practically inconceivable” in the sense that it is practically necessary for beings like us, but it is not therefore “absolutely inconceivable” (Strawson, 2008a, p. 14). Having the concept is conditional—but only on maximally generic conditions of human life.²¹⁵ If we did not have the concept of responsibility and some practice of blame, it is very “doubtful whether we should have anything that *we* could find intelligible as a system of human relationships, as human society” (Strawson, 2008a, p. 26, original emphasis).

The explanation also makes explicit the sense in which the concept of responsibility is *non-historical*. We arrive at a generic version of the practice of blame and see the fundamental character of the concept of responsibility being formed already from *within* our fictional State-of-Nature. The practical necessity of the concept of responsibility *precedes* the historical conditions of any *actual* human form of life.²¹⁶

²¹⁵ “I think it is very, very unlikely” that we completely “refrain from moral reaction to other people” (Strawson et al., 2008, p. 95).

²¹⁶ There is a weaker version of this claim. We may distinguish between two contrasting ideas of a concept being necessary for any human society. Either the concept is non-historical in the sense that it is *pan-historical*—i.e., in any *actual* human society, there is some conception of responsibility—or it is non-historical in a sense that goes beyond that; to put it in a way that only the contrast allows, in the sense that it is *genuinely non-historical*. Either may be

The pragmatic genealogy also makes explicit what *form* the non-historical concept has—“the fundamental character” of the concept of responsibility. Since the practice of blame is self-effacingly functional, the concept of responsibility that sustains this practice must be *backward-looking*. This, it turns out, is not a contingent feature of *our* conception, but a necessary feature of any conception of responsibility that can be sufficiently stable to actually be a conception of responsibility. The genealogy provides an explanation of the autonomous (or non-instrumental) nature of the practice in terms of the functional dependence of the practice on this autonomy. This is a very important point, for several reasons:

A general opinion is that Strawson’s ‘Freedom and Resentment’, while it professes to reconcile the instrumentalist optimist and the metaphysical pessimist, in fact aims to leave them both behind. In particular, while instrumentalism about responsibility was popular around the paper’s publication (e.g., Nowell-Smith, 1948; Schlick, 1939, ch. 7; Smart, 1961),²¹⁷ The paper is generally considered as having furthered, or at least been expressive of, the loss of interest in that tradition since.²¹⁸ As Stephen Darwall articulates this thought:

Strawson argued that social desirability is not a reason of “the right sort” for practices of moral responsibility “as we understand them.” When we seek to hold people accountable, what matters is not whether doing so is desirable, either in a particular case or in general, but whether the person’s conduct is

sufficient for capturing the thought that for any actual human society, it is a precondition on its existence that it has *some* conception of responsibility, some development of that most fundamental concept. The weaker sense might be enough. However, I do think that the *fictiveness* of the State-of-Nature condition represent the concept as not merely part of any actual conceptual scheme—not merely *pan*-historical—but allows us to see how the concept, in its fundamental form, really is not an historical artefact at all—that it is *non*-historical. The pragmatic genealogy allows a way of making naturalistic sense of that rather lofty thought. Unlike the (merely) *pan*-historical sense, on the genuinely non-historical sense, the concept of responsibility is conceived as antecedent the actuality of any actual human society; the generic version of the practice is practically necessary already in the highly idealised State-of-Nature condition. Perhaps this difference should not be insisted on here; it might well be that what Strawson means with “non-historical” is (on the present distinction) “*pan*-historical”. Either way, the claim is, the pragmatic genealogy provides a way of articulating, in naturalistic terms, the idea that a concept is, in its fundamental form, without history.

²¹⁷ Strawson (2008a, n. 1) mentions Nowell-Smith (1948) as an example of the instrumentalist optimist.

²¹⁸ See for example Darwall (2006, p. 15f.), Watson (2004), Wallace (1994), Shoemaker (2020, p. 225., esp.) and Russell (2017b).

culpable and we have the authority to bring him to account. Desirability is a reason of the wrong kind to warrant the attitudes and actions in which holding someone responsible consists in their own terms. (2006, p. 15)

Without contradicting the historical picture or even this interpretative point, many see Strawson's approach (even if underdeveloped) as a *functionalist* or *expressivist* account of responsibility (e.g., Bennett, 1980).²¹⁹ Today, communicative (e.g., Macnamara, 2015a, 2015b; Watson, 2004, 2011) or conversationalist (McKenna, 2012) accounts have become popular, and they typically take their influence from Strawson.²²⁰ But also instrumentalist or influenceability accounts have become increasingly popular (e.g., Arneson, 2003; Barrett, 2020; Jefferson, 2019; McGeer, 2012, 2013, 2019; Milam, 2021; Vargas, 2013). This revival of instrumentalism is in no small part due to these authors having incorporated several insights from Strawson's account (cf., Barrett, 2020; Jefferson, 2019; McGeer, 2019; Miller, 2014b; Vargas, 2004).²²¹ This recruitment of "Strawsonian ideas" have even led some to propose—as an interpretation—that Strawson himself in fact advanced an (indirect) instrumentalist account of responsibility (McGeer, 2014; Miller, 2014a), directly opposing the established view of Strawson's approach and its legacy.²²²

Given this state of affairs, explaining the conceptual practice of blame as self-effacingly functional may itself strike a recognisably Strawsonian tone—that of seeking to reconcile the disputants. One such potential comes from the fact that the insight that the fundamental character of the practice of blame is backward-looking is itself an insight from considering the efficacy of blame. It turns out that our practice of blame has this character not merely coincidentally or contingently—but has it *for functional reasons*: the very functionality of the practice depends on the practice being autonomous, i.e., not dependent on the functionality of the practice in any particular instance of blame. The function of blame is not dismissed as irrelevant: it is *explanatorily*

²¹⁹ See also, for example, Beglin (2018).

²²⁰ See, e.g., Macnamara (2015b), Helm (2012), Smith (2012), Fricker (2016), and Shoemaker (2021).

²²¹ Several instrumentalists explicitly appeal to Strawson's insight of the centrality of the reactive attitudes in presenting their renewed instrumentalism. Even Arneson with his "Smart theory" does (see 2003, p. 239 n. 12, p. 240 n. 13).

²²² However, read carefully, each involves certain qualifications. McGeer's reading saves Strawson from "naturalistic complacency"—but it is reasonable to ask, when considering this as an interpretation, is that Strawson's worry or McGeer's worry?

basic. But the function is irrelevant for the practice in the sense that blame is *necessarily* guided by autonomous (non-instrumental) reasons. However (to take one more turn), this necessity is itself understood in terms of the functionality of the practice. This serves as a move towards reconciliation in that it provides a designated space for each of the insights of the opposing parties. (Of course, it might just be that this, as any “move towards reconciliation [...] is likely to seem wrongheaded to everyone” (Strawson, 2008a, p. 2).)

Relatedly, the present reading allows us to make sense of that last passage of ‘Freedom and Resentment’, which might otherwise seem rather perplexing:

If we sufficiently, that is *radically*, modify the view of the optimist, his view is the right one. It is far from wrong to emphasize the efficacy of all those practices which express or manifest our moral attitudes, in regulating behaviour in ways considered desirable; or to add that when certain of our beliefs about the efficacy of some of these practices turns out to be false, then we may have good reason for dropping or modifying those practices. What *is* wrong is to forget that these practices, and their reception, the reactions to them, really *are* expressions of our moral attitudes and not merely devices we calculatingly employ for regulative purposes. Our practices do not merely exploit our natures, they express them. Indeed, the very understanding of the kind of efficacy these expressions of our attitudes have turns on our remembering this. When we do remember this, and modify the optimist’s position accordingly, we simultaneously correct its conceptual deficiencies and ward off the dangers it seems to entail, without recourse to the obscure and panicky metaphysics of libertarianism. (Strawson, 2008a, p. 27, original emphasis)

While the indirect instrumentalist interpretation of Strawson, driven by the “want to know what justifies these attitudes and practice *as a whole*” (McGeer, 2014, p. 90, original emphasis), ascribes to him *an external justification* in terms of efficacy—a thought we should recognise as antithetical to Strawson’s approach—the pragmatic genealogy does not saddle Strawson with such a claim. On a pragmatic genealogy of blame, the efficacy is shown to be debarred from *the sphere of justification*. It is relevant only for *the sphere of explanation*. (We will return to this contrast shortly.)

Finally, and most importantly, that the fundamental character of the concept of responsibility is essentially backward-looking bears directly on the implication of Minimal Optimism for responsibility scepticism. Without this insight, it might have seemed that the Minimal Optimist idea—that the concept of responsibility is inescapable without any particular conception of responsibility being inescapable—leaves it completely open for the sceptic to simply reconceptualise their proposal without a loss. The responsibility sceptic

may simply accept that we need *some* conception, while nevertheless insisting that the conception we currently have is unjustified. However, the sceptic-friendly conception of responsibility they wish to replace our conception with is typically a purely instrumental one. Omitting notions of desert, fitting shame, and the like, their conception is one stripped of the backward-looking features of blame and responsibility.

By supplementing Strawson's approach with the pragmatic genealogy above, Minimal Optimism turns out to be more challenging for the sceptic than it might otherwise have seemed. The very *concept* of responsibility essentially (since necessarily) has a backward-looking character. Any *conception* of responsibility must be a "development of, or from" *that*. Strawson's approach, on the Minimal Optimist Reading, is thus anti-sceptical in a way that speaks to our current dialectical situation.

VI. Explanation, Justification, and Vindication

As we've seen, Strawson recognises the importance of the distinction between *justification* and *explanation*. The pragmatic genealogy of blame makes essential use of this distinction: while the function of the practice pertains to *the sphere of explanation*, the considerations in *the sphere of justification*—i.e., the considerations that serve as reasons for applying the concept in any given case—are not conditional on the function of the practice. The genealogy furthermore provides an explanation of *why* the sphere of justification is autonomous in this sense: for functional reasons. It is part of the structure of a self-effacingly functional practice—such as blame and, by extension, responsibility—that these two spheres are distinct.

This explanation, just because it is *not* a justification, is not therefore normatively inert or evaluatively impotent: the genealogy can *vindicate* our living by a particular concept.²²³ We may distinguish between three kinds of vindication that a pragmatic genealogy is apt to achieve: a *negative*, a *naturalistic*, and a *pragmatic* vindication (Queloz, 2021a, pp. 98, 178).

Negative vindication is achieved in so far as the genealogy does not draw on or reveal anything that would be incompatible with living by that concept. This is the inverse of debunking explanations: the genealogy is vindictory in this sense in so far as the genealogy is *not* a debunking explanation of the concept.

²²³ Glock (2022) ascribes, despite Strawson's contestations, a vindictory spirit to Strawson's descriptive metaphysics. However, this is not the kind of vindication at issue here.

Naturalistic vindication is achieved in so far as we can make naturalistic sense of the concept that might otherwise have seemed mysterious or ethereal. This is particularly pertinent for the concept of responsibility, intimately connected to various ideas of free will as it is. Many harbour the suspicion that the practice of blame depends on a loitering concept of an enchanted time, a concept of responsibility that presupposes such metaphysically fantastic ideas as that of being *causa sui* (cf., Nietzsche, 2014, §21). By understanding the concept rather in terms of human needs that we have reasons to acknowledge independently of that concept, we have made naturalistic sense of the concept.

Pragmatic vindication is achieved by showing the people that are engaged in the practice that it is *instrumentally rational* for them to be engaged in the practice. If they reflect on why they live by the concepts they live by, they are shown that it is *practically rational* for them, given the way they are and the kinds of situations they face, to continue to live in this way. Importantly, this does not mean that they are *justified* in living by the concept nor that they are justified in any particular application of the concept. Because it is instrumentally rational for beings like us to, for example, have a conception of responsibility, it follows, neither, that the practice as a whole is justified, nor that anyone is justified in holding someone responsible. For one thing, whether anyone is justified in holding someone responsible will depend on whether the person satisfies the relevant criteria for being responsible (the rules of the practice); and these criteria are (necessarily—as our genealogy shows) not spelled out in terms of the instrumental value of living by a conception of responsibility. Nevertheless, the genealogy has evaluative import for those people already within the practice. We who have a conception of responsibility, who hold responsible and take responsibility, who already have some criteria for being responsible, we may have our *confidence* in continuing to live by this concept bolstered. The genealogy provides the concept-users with reasons to prefer this practice over any alternative, including the alternative of abandoning the practice; that is, even if they did not and could not chose to engage in the practice, they are given reasons to favour the practice. In light of the pragmatic genealogy, we may thus engage in the practice with reasonable confidence.

The potential to rationally and evaluatively inform how we conceive of our engagement in these practices is a virtue of the method of pragmatic genealogy, specifically when serving as Strawson's missing answer. The philosophical importance of the explanatory task, according to Strawson, is to further our understanding of the conceptual scheme and to assist the philosopher in accurately describing our conceptual scheme. But the recurrent juxtaposing of, on the one hand, the limits of justification and, on the other, the possibility of

explanation, suggests something more—that there is a point to providing *explanation in the place of justification*. The evaluative force of a pragmatic genealogy—to affect our confidence in engaging in the practice—makes sense of (indeed, buttresses) this replacement of justification with explanation.

In chapter 4, we considered the Reversal Move, the idea that the practice is somehow prior to the concept. We found that, what is lacking in the accounts considered if they are to make good on Strawson's Promise is a naturalistic explanation that explains why the concept of responsibility is such that the truth of determinism is irrelevant. This seems to be what we want if we are to say that something like the Reversal Move makes good on Strawson's Promise. The here proposed answer to the unanswered question of 'Freedom and Resentment' is of that sort. In the next section, we'll see how this naturalistic explanation buttresses the Inescapability Claim, rendering that anti-sceptical response more plausible. Before this, in relation to the vindicatory force of the pragmatic genealogy, it is worth highlighting a contrast to other naturalistic explanations of Strawsonian decent.

A shortcoming of these other Strawsonian explanations is that they neglect the issue of the *authority* of the practice. Even if we admit that what it means to be responsible is determined by the practice of holding responsible, it is a legitimate question to ask, 'Why do we engage in this practice?'. If we recognise the fact that the practice (somehow) determines the concept, the thought that *we did not make the ideas we live by* is not rendered any less acute—to the contrary.²²⁴ To simply rely on the *wisdom of the ages*—to trust that people past have held each other responsible in suitable ways, and so trust that the concept of responsibility that we have inherited is also adequate—appears as an inadequate response to a legitimate worry.

Naturalistic explanations of responsibility usually make *some* reference to our basic needs (or 'basic concerns') (e.g., Beglin, 2018; Bengtson, 2019; Watson, 2008, p. 117).²²⁵ However, while the connection may be there, the evaluative import of understanding our concepts in terms of our needs is not

²²⁴ All the more serious, that is, compared to the salient alternative, of a practice-independent notion of responsibility being tracked by (or that should be tracked by) our practice.

²²⁵ Also Shoemaker makes some reference to what is known as the 'basic concern' (cf., Watson, 2008) when he says that "fundamental to our nature is a kind of sociality, and given our sociality, it matters greatly to us what intentions and attitudes others have toward us" (D. Shoemaker, 2017, p. 482). I worry that such formulations reveal that the 'basic concern' is not understood in a basic enough way; it appears to be a rather thick notion, tied up with "our social sentimental nature", our "sociality", and "our humanity". The worry is that too much is built into it for the account to have either explanatory or dialectical force. The pragmatic genealogy reaches further down, so to speak; to more neutral ground.

made clear. Consider again the challenge from the possibility of collective mistake: The critic asks the Reversal Theorist: If we were to hold, say, very young children responsible, would very young children then *be* responsible? This is absurd: very young children are *not* responsible. Therefore—so the objection goes—accounts that take the practice to determine the concept must be mistaken (Fischer & Ravizza, 1993, p. 18; Todd, 2016). What we should first of all say in response to this is that a *different* practice would entail a *different* conception of responsibility and, thus, very young children would “be responsible” only according to *that* conception of responsibility, but they would not be responsible in *our* sense of the term (Bengtson, 2019; De Mesel, 2022a, p. 1909; also, De Mesel & Heyndels, 2019).²²⁶ Radically different practices might mean that we would “*call* children ‘morally responsible’, but ‘responsible’ would express a concept[ion] that is different from ours” (De Mesel, 2022a, p. 1909, original emphasis).²²⁷ This is correct, as far as it goes.

But even if *our* practice determines *our* conception of responsibility, whence *the authority* of our conception? If we had a different conception of responsibility, would that conception then necessarily or automatically be (just as) authoritative? We need not stay silent on this issue. Explaining concept-use in terms of practical needs can, in so far as these are needs that we acknowledge, pragmatically vindicate engaging in the practice. We need not merely think, ‘We *just happen to* be this way, so we do these things and think in these ways; if we were different, we would just do differently and think differently’.²²⁸ The pragmatic genealogy shows us, beyond this, that to do and to think in the way we do is *useful* given the way we are. The practice is vindicated to the extent that our practice is, given our needs, preferable—we may continue in this way with reasonable confidence.

To break the silence on this point is important for the rehabilitation of Strawson’s approach. It counters the claim that Strawson’s view is “relentlessly naturalistic” (Coates, 2017, p. 804), or ‘naturalistically

²²⁶ Beglin (2018, p. 622) considers the issue in counterfactual terms: “it is not obvious that rational agents who relate to each other in fundamentally different ways [...] would or should operate with the same conceptual equipment as we do”. Plausibly, we should understand Beglin as reaching for the same point: that a different practice would constitute a different concept(ion) of responsibility (De Mesel, 2022a, p. 1909).

²²⁷ As we are using the terms, it would express our *conception*.

²²⁸ Beglin might be touching on this point, however, when he remarks that “it isn’t obvious that rational agents who relate to each other in fundamentally different ways, who have fundamentally different evaluative standpoints, which are constituted by fundamentally different basic concerns, would *or should* operate with the same conceptual equipment as we do” (2018, pp. 623, emphasis added).

complacent' (McGeer, 2014, pp. 88, 89), or problematically "conservative" (McKenna & Russell, 2008a, p. 10). By Strawson's lights, we have to "begin at the beginning. And not [...] try to go further back" (cf., Strawson, 2008c, p. 19; Wittgenstein, 1969, §471). But this does not mean, as Kant complained, that our conceptual scheme "must be accepted merely on *faith*" (Kant, 1781, B xi; Strawson, 2008c, p. 19). The pragmatic genealogy starts from the recognition that "[w]e did not make the ideas we live by" (Queloz, 2021a, p. 1) and, from that recognition, it seeks to understand why we have the concepts we have. If the genealogy is vindictory, while we have still not ourselves made the ideas we live by, in recognising what these ideas do for us, our confidence in living by them is strengthened.

A further contrast between existent Reversal Theses, on the one hand, and the method of pragmatic genealogy, on the other, is that the latter is better suited to account for the autonomous or backward-looking character of responsibility. While the Reversal Theses considered (De Mesel, 2022a; D. Shoemaker, 2017, 2022) are able to accommodate the traditional idea that it is only appropriate to hold someone responsible if they are responsible, they do not explain why we have this idea. The pragmatic genealogy does not only explain why we have it but also why we *need* to have it. This adds an anti-sceptical edge to this account that the others lack.

The Meaning-Based Account, as the pragmatic genealogy above, distinguishes between the sphere of explanation and the sphere of justification. This allows the Meaning-Based Account to explain our concept of responsibility in terms of the practice, while accommodating the traditional idea in the sphere of justification. While the concept of responsibility is *constitutively dependent* on the practice of responsibility, to hold someone responsible is *justificatorily dependent* on them being responsible.

A virtue of the pragmatic genealogy is that it elucidates what the relationship is between the sphere of justification and the sphere of explanation beyond the fact that they are distinct spheres. The Meaning-Based Account does not tell us whether the distinctiveness of these two spheres is contingent or necessary—and, if necessary, why it is. By understanding the practice of blame as self-effacingly functional, the pragmatic genealogy shows that the reasons in the sphere of justification are *necessarily* distinct from those in the sphere of explanation. They are necessarily distinct because the functionality of the practice itself depends on them being distinct. Though both the Meaning-Based Account and the pragmatic genealogy can accommodate for the justificatory autonomy of the practice, the former *merely accommodates* it, while the latter *accounts* for it. As we saw in the last section, this insight of the pragmatic genealogy is an important one for several reasons.

VII. The Fundamental Irrelevance of Determinism

This answer to the unanswered question of ‘Freedom and Resentment’ serves to buttress the Inescapability Claim. What the pragmatic genealogy stands to show is that determinism is *fundamentally* irrelevant for responsibility.

For one thing, the concept of responsibility that is practically necessary for any human society simply in virtue of some very basic needs and situations is plausibly relatively metaphysically undemanding. Likewise, the generic practice of blame that is practically necessary simply in virtue of very general facts about human society is plausibly not a very metaphysically demanding practice. In order to serve only these basic needs, what is needed is not something metaphysically demanding; plausibly, not something that requires the falsity of determinism.

What determinism is said to be irrelevant for is only the concept of responsibility and the practice of blame *in their fundamental form*. What is “for us as we are, practically inconceivable” (Strawson, 2008a, p. 12) is that we would not have *the concept* of responsibility and the *generic* practice of blame. This does not mean—as the Minimal Optimist is apt to stress—that any *particular conception* of responsibility is inescapable for human beings. Our human nature—the most basic facts about us, our needs and situations—*underdetermines* our conceptual scheme and ethical practices. Nevertheless, given that we cannot have either the concept of responsibility or the practice of blame in their non-historical character, it follows that to have *some* conception of responsibility and *some* practice of blame is practically necessary for beings like us. While determinism might be relevant for some of these particular conceptions that we might have, what it is irrelevant for is whether we have *a* conception of responsibility.

A further sense in which determinism is fundamentally irrelevant is that the needs and situations that render the concept of responsibility practically necessary are *basic* needs and *basic* situations. While the last point underscored the openness and contingency of our actual practices, this point underscores the limits of this openness. Even though our conceptual scheme and ethical practices are underdetermined, they are nevertheless, in some fundamental respects, determined. Our framework commitment to the concept of responsibility “is a natural fact, something as deeply rooted in our natures as our existence as social beings” (Strawson, 2008c, p. 26). It is inescapable for us because it is practically necessary already on the assumption only of very basic needs of any human society. These are needs we have whether or

not determinism is true.²²⁹ In that sense too, determinism is fundamentally irrelevant for the concept of responsibility.

VIII. Conclusion

In this chapter, I've argued that the method of pragmatic genealogy is well suited to provide the answer to the unanswered question of 'Freedom and Resentment'. Even if the reader does not find the method of pragmatic genealogy appealing, the subtext reading of Strawson that emerges from the congruity, which I have argued for, of that method with Strawson's approach may still be illuminating.

The answer to the unanswered question of 'Freedom and Resentment' here proposed offers reconciliation of opposing parties in the traditional and the contemporary discussion of free will, as well as of opposing parties regarding the interpretation of Strawson's approach.

Primarily, the pragmatic genealogy buttresses and explicates the Inescapability Claim. In rendering us having a conception of responsibility intelligible—showing that beings like us would have practical reasons to have something like the conceptual scheme that we in fact have—the pragmatic genealogy *vindicates* the concept of responsibility. For beings like us, we see, the concept of responsibility is inescapable, fundamentally, in the sense that it is *indispensable*.²³⁰

In this light, the Inescapability Claim seems neither "relentlessly naturalistic" (Coates, 2017, p. 804), naturalistically complacent (McGeer, 2014, pp. 88, 89), nor problematically "conservative" (McKenna & Russell, 2008a, p. 10). The criticism that the Inescapability Claim is empirically or naturalistically implausible seems misplaced. On the Minimal Optimist understanding, in light of the kind of naturalistic explanation that our pragmatic genealogy exemplifies, the grounds for dismissing the Inescapability Claim start to appear rather feeble.

Rendering worries of complacency off tune, this answer roots Strawson's Promise and the anti-sceptical stance of Strawson's way with responsibility in a naturalistic and realistic conception of the human being. The answer here provided goes at least some way towards vindicating Strawson's sense that

²²⁹ Plausibly, these are needs that we would have also if we fully grasped the truth of determinism, which a Spinozistic sceptic would argue that we have not.

²³⁰ I thank Matthieu Queloz for this formulation.

reflection on our human nature can reveal that we have “all that we actually need, or can hope for, in the way of responsibility and freedom” (Strawson et al., 2008, p. 93f.).

7. Minimal Optimism & Strawson's Optimism

I. Introduction

In the last two chapters, I've argued that Strawson's view is a Minimal Optimism, and have explicated and substantiated that view as Strawson's view.

This chapter seeks to go beyond Strawson's Minimal Optimism. By picking up (in section II) on one of the standard criticisms against the Inescapability Claim—the inadequacy criticism—a further sense in which Minimal Optimism makes good on Strawson's Promise is to be made explicit. Minimal Optimism serves to thwart the tendency to privilege our own conception of responsibility as *the* conception, thus averting the risk that scepticism about a particular conception of responsibility becomes scepticism about the concept of responsibility. This highlights a critical edge to Minimal Optimism, to change the terms of discussion and bring us beyond the traditional worries.

There is an issue regarding 'the spirit' of this approach—that is, what concerns that are driving the project, what the aims and ambitions of the approach are. The problem is that the approach presented does not seem to share Strawson's spirit. A formal solution is presented (in section III), but we will return to the issue of spirit towards the end (in section V).

A further respect in which we'll go beyond Strawson's Minimal Optimism in this chapter is that we will consider, and reconsider, Strawson's optimism about our particular conception of responsibility (in section IV). The most instructive piece of text for illuminating this matter is the exchange between Strawson (1980, pp. 264-266) and Bennett (1980). We will entertain a few different readings of Strawson's response: that his optimism about our conception is mistakenly grounded on his Minimal Optimism; that he is in fact *only* a Minimal Optimist, and takes this to be sufficient; and, last, that he takes the idea behind his Minimal Optimism to be analytically useful for defending an optimism about our particular conception. The latter is the most promising, and the most intriguing, reading.

Towards the end (in section V), we consider some reasons for thinking that, while Strawson was content to declare certain differences in outlooks mere differences in temperament, we should take seriously the thought that his approach is more critical to our actual ethical outlook than he says it is. In light, not of what he says he says, but of what he actually says, his approach indeed seems to be more critical—indeed, more *revisionary*—than he himself suggests.

II. On Inadequacy and Historical Sense

One of the standard challenges against the Inescapability Claim, which we have not considered yet, is that it is *inadequate*. Contrary to the implausibility criticism, which holds that the claim is too strong, the inadequacy criticism holds that the claim is not strong enough. The key point of *Minimal Optimism* consists in limiting the Inescapability Claim. However, is Minimal Optimism *too* minimal?

With respect to Strawson's approach, there are two different versions of the inadequacy criticism, which may be distinguished in terms of what the approach is purportedly inadequate *for*:

- (a) The inadequate-*for-defending-an-ethical-outlook* criticism (Smilansky, 2008, pp. 242-244).
- (b) The inadequate-*for-rejecting-scepticism* criticism (Russell, 2017b, 2017d).

As they pertain to the present reading of Strawson, both of these challenges are to be defused by rejecting an illicit privileging of any particular conception of responsibility. They may both be rejected in the same way because, by the Minimal Optimist's lights, they come to the same thing. Let's present the criticisms in a bit more detail in order to see why.

First, the inadequate-*for-defending-an-ethical-outlook* criticism: Saul Smilansky (2008, pp. 242-244) argues that Strawson's naturalistic approach is "inadequate" because it does not provide a sufficient ground for a *certain* conception of responsibility, and thus cannot ground certain ideas of justice and respect.²³¹ These are ideas that we are said to care about; ideas which Smilansky takes to "represent moral progress" (see also, Smilansky, 2000).

²³¹ He claims, without further explanation, that it is inadequate "even on its own naturalistic terms" (2008, p. 242).

The criticism is that Strawson's naturalism is inadequate for defending *this particular* ethical outlook (and we might add, for the same reasons, that it is inadequate for defending *any* particular ethical outlook).

Second, the inadequate-for-rejecting-scepticism criticism: Russell (2017d) poses this as one horn of a dilemma. He distinguishes between two different "kinds or modes" of naturalism and pessimism. While the one naturalism is (naturalistically) implausible, the more limited and more plausible is inadequate for rejecting a sensible kind of pessimism/scepticism.²³² While Russell's characterisation of Strawson's view is in the terms favoured by the Standard Reading—and for that reason not to be considered directly—an analogous suspicion may arise just as well with respect to the Minimal Optimist Reading. Let's transpose the challenge, distinguishing between two kinds or, better, levels of optimism and scepticism respectively:

²³² *Type-naturalism* holds that reactive attitudes are inescapable at the level of our human "disposition or liability to reactive attitudes" (2017d, p. 39, emphasis added). Against the pessimist/sceptic, it holds that our "susceptibility" to a particular "*species* or *type* of emotion" does not require an external 'rational' justification. (It is noteworthy that there are *two* abstractions, not one, in type-naturalism: (i) to *types* or *species* of emotions—concerned with, for example, (the 'natural emotion') *anger* rather than (the 'cognitively sharpened') *resentment*—and (ii) to *disposition* to emote rather than actually emoting. The contrast between type- and token-naturalism is not one dimensional.) *Token-naturalism*, on the other hand, holds that we inescapably "*feel* or *experience*" or "*entertain*" (2017d, p. 37, emphasis added) particular *tokens* or *episodes* of the reactive attitudes. (Naturalism, on Russell's reading, is committed to the claim that we are liable, disposed or prone to the reactive attitudes, *or* feel, experience, or entertain tokens of the reactive attitudes "whatever reasons suggests to us" (2017d, p. 37). At the level of *type*, according to Russell, this claim just is the repudiation of the need for an "external 'rational' justification". At the *token* level, however, the claim is that "[i]f reason were to point us in this direction [...] we would be constitutionally incapable of following its lead" (2017d, p. 37). Here, "constitutionally incapable" means psychologically incapable.) *Type-pessimism* demands a rational justification for our mere *proneness* or *susceptibility* to reactive attitudes but finds none (at least, if determinism is true). *Token-pessimism*, on the other hand, requires that any *instance* or *episode* of the reactive attitudes requires a rational justification. It holds that whether we "entertain or engage" (2017d, p. 39) an episode or token of a reactive attitude in a given case is, or should be, dependent on whether the circumstances are such that we are justified in entertaining or engaging the emotion under these circumstances. In addition to this demand for justification, the token-pessimist holds that we are never in such circumstances (at least, if determinism is true). Russell (2017d, p. 37) rightly concludes that token-naturalism is naturalistically implausible—and that it is not Strawson's view. The problem, on Russell's dilemma, is that this leaves Strawson with the inadequate response of type-naturalism. While type-naturalism is incompatible with *type-pessimism*, it leaves *token-pessimism* perfectly intact. Merely claiming that we are inescapably *prone* to reactive attitudes does not address the sceptical worry that circumstances are never such that we are justified in actually having reactive attitudes.

- (I) *Concept-optimism*: Optimism about the concept of responsibility.
- (II) *Conception-optimism*: Optimism about a conception of responsibility.
- (III) *Concept-scepticism*: Scepticism about the concept of responsibility.
- (IV) *Conception-scepticism*: Scepticism about a conception of responsibility.

Minimal Optimism is concept-optimism, and hence incompatible with concept-scepticism. However, so this version of the objection goes, concept-optimism is perfectly compatible with conception-scepticism; therefore, Minimal Optimism is inadequate for rejecting a sensible kind of scepticism.

To be clear, Minimal Optimism is only compatible with conception-scepticism if this is understood as scepticism about *a particular conception* of responsibility. The inescapability of the non-historical concept of responsibility means that to have *some* (actual, socio-historically particular) conception of responsibility is also inescapable, even if *no one* particular conception is in that sense inescapable. Minimal Optimism is thus not compatible with a scepticism about any or every conception of responsibility, as such general conception-scepticism is tantamount to concept-scepticism.

Once the second inadequacy challenge has been so redescribed, it reduces to the first. Whether circumstances are such that the application of the concept of responsibility on some ethical outlook is justified will depend on what the criteria for application are on the particular *conception* of responsibility of that outlook. The criticism is simply that Minimal Optimism is silent on whether any particular conception is defensible or misguided.

That Minimal Optimism does not defend or ground any particular conception of responsibility is in fact something it should insist on. While stressing the common core of human thought is meant to constrain conceptual variation, thus fending off unqualified relativism or historicism, Minimal Optimism likewise opposes the tendency, as Nietzsche (1986, §2) puts it, to “think of ‘man’ as an *aeterna veritas*, as something that remains constant in the midst of all flux”.²³³ It is part of the point of stressing that there is a fundamental core of human thinking which has no history that it helps us see that the untenability of a particular conception should not be cause for scepticism about responsibility, but *only* scepticism about that particular ethical outlook and its particular conception of responsibility. The delimitation of the Inescapability Claim is in recognition of the fact “that man has become”

²³³ See also Nietzsche (1889/1997, Reason §1; 2003, I §2, II §4).

(Nietzsche, 1986, §2)—i.e., that our ideas have a history. To distinguish some features of our conceptual scheme as constituting the “minimum structure” or “core of human thought” that is inherent in any human conceptual scheme is meant to avert the privileging of *some* conception as *the* conception—i.e., to think of it as expressing the concept in its “pure” (non-historical, non-contingent) form.²³⁴ This is to show an historical sense.

As noted, Strawson (2008a, p. 26) suggests that in so far as his own “descriptions of human attitudes have reflected local and temporary features of our own culture”, to that extent it is imperfect (cf., Russell, 2024, p. 247f.). This does not mean that our conception falls short because it does not adequately represent ‘the concept’. To the contrary, it is meant as a correction precisely against this kind of privileging of our own conception.²³⁵ It is in the fact that it is not advancing any particular conception, but emphasises the common core of each such conception, that Minimal Optimism’s full anti-sceptical force may be seen.

This point is important, because to make our own conception out as *the* conception encourages *nihilism* about responsibility: reasonable scepticism about our own conception easily becomes scepticism about any conception, as the sceptic remains committed to the idea that our conception is *the* conception, or the only one that there could have been; and hence thinks that, nothing, no other conception, can replace it.²³⁶ Indeed, the tendency to so privilege an outlook is itself particularly prominent in a particular ethical outlook, which is in some measure part of our outlook.²³⁷ In this light, Minimal Optimism emerges with more of an edge than it might at first have seemed to have.²³⁸

²³⁴ This is part of Williams’s strategy in both *Shame and Necessity* (1993) and *Truth and Truthfulness* (2002; cf., 2014, p. 407).

²³⁵ Cf., “That idea—‘the wrong-doer deserves punishment *because* he might have acted otherwise,’ in spite of the fact that it is nowadays so cheap, obvious, natural, and inevitable, and that it has had to serve as an illustration of the way in which sentiments of justice appeared on earth, is in point of fact an exceedingly late, and even refined human judgement and inference; the placing of this idea back at the beginning of the world is simply a clumsy violation of the principles of primitive psychology” (Nietzsche, 2003, II §4).

²³⁶ The tendency to privilege our own conception in this way is the result, from this perspective, of what Williams calls ‘the purity of morality’ (1985, pp. 216-218); a central point for Nietzsche (see, Clark, 2015; Queloz, 2022b; Queloz & Cueni, 2019).

²³⁷ That is, what Williams (1985, ch. 10) calls ‘the morality system’.

²³⁸ Russell (2017c) makes roughly the same point when arguing against Wallace’s (1994) ‘narrower’ construal of the reactive attitudes in favour of a ‘broader’ construal. I say “roughly”, because the distinction between a narrower and a broader construal of the reactive attitudes is distinct from the distinction, central to the Minimal Optimist Reading, between the concept and any conception of responsibility. The two should not be conflated. The

This is the final sense in which Minimal Optimism may be said to make good on Strawson's Promise. This sense concerns Strawson's Promise not, in the first instance, with respect to its claim that determinism is irrelevant for responsibility but, rather, with respect to the sense that it brings us beyond the traditional debate. Recognising that we necessarily have *some* conception of responsibility, scepticism about responsibility is contained to the particular conception of responsibility that provokes it; the sceptic's challenge, thus contained, becomes the (moderately) revisionistic one that we should have *another* conception of responsibility. To reject the concept of responsibility

concept/conception distinction is in fact an interpretatively more appropriate distinction. To see how, consider Russell's broad/narrow distinction and his criticism of Wallace on the basis of it. On Wallace's (1994) account, we should adopt a *narrow* construal of the reactive attitudes (cf., De Mesel & Cuypers, 2023; Sars, 2022b). *Moral* responsibility is to be understood in terms of *moral reactive attitudes*. The moral reactive attitudes, on Wallace's account, belong to "a special sphere within our ethical concepts", namely, that of "moral rightness and wrongness" (Wallace, 1994, p. 63). This distinguishes moral reactive attitudes from other, non-reactive moral sentiments, such as shame, gratitude, and admiration (Wallace, 1994, p. 37). The latter class of attitudes are not "based" on the same kind of moral values; they are not elicited strictly by the violation of moral obligations. Against Wallace's construal, Russell (2017c) argues, for one, that the asymmetrical view of moral responsibility—connected exclusively with the *negative* reactive attitudes of resentment, indignation, and guilt; i.e., with responses to violations of obligations—truncates our understanding of responsibility, and effectively ignores "the constructive role of reactive attitudes" as directed to worthy and admirable actions. Russell also argues that the narrow construal raises a "conceptual barrier" between our form of ethical life and others, historically and culturally distant from us. These others have, on the narrow construal, at best an "analogous" practice to "the" (i.e., *our*) moral responsibility practice. Such conceptual privileging of our form of ethical life is problematic, Russell argues, because it distorts self-understanding, presenting our conception as more distinct from other conceptions than it is—as 'conceptually isolated'—and also as if more coherent than it is. Narrowing thereby impoverishes our resources for critical self-evaluation. Russell's argument for a broader construal overlaps considerably with the argument against privileging our conception discussed above. However, as a matter of interpretation, the concept/conception distinction is more apt than the broader/narrower distinction because it is, in the following two respects, relevantly sharper. The problem with narrowing is *multi-dimensional*: narrowing produces, as Williams explains, a "damaging isolation *both* from other ethical ideas *and* from the rest of life more generally" (1995d, p. 203, emphasis added). A construal is broader if it is less isolating in either of these two dimensions. While conceptual isolation of our ethical outlook from other ethical outlooks is antithetical to Strawson's principal concern, the distinctiveness of the moral and the non-moral is less clearly a Strawsonian concern. This is the first respect in which a broader construal of the reactive attitudes is not sharp enough with respect to the interpretative issue. Second, while the broad/narrow distinction is *gradual*, the concept/conception distinction is categorical. And again, interpretatively, this is better. The distinction should be categorical: it is between "the core of human thinking which has no history" and the "developments of, or from", that core. The suggestion of a broader construal is not as well-anchored in Strawson's approach as the concept/conception distinction.

itself and, with it, any and all conceptions because *our* conception of responsibility, with its associated socio-historically local ideas of justice and freedom, is unsustainable appears, in this light, as an act of bad faith.

III. Minimal Optimism and Strawson's Spirit

The point of the last section does not sound like a point Strawson would make. The reader may begin to wonder, though Minimal Optimism may be true to the letter, whether it has strayed too far from the *spirit* of Strawson's approach.

We might think that there is a tension in the Minimal Optimist Reading because Strawson himself seems more optimistic than this—he does not seem to be bothered by *any* scepticism about responsibility, not even scepticism only about our conception of responsibility. If Strawson is in fact optimistic about our conception, this explains why he would not be inclined to make the point of the last section. This does not mean, however, that Minimal Optimism is too minimal to be Strawson's view. Above, we distinguished between two levels of optimism.

(I) Optimism about the concept of responsibility.

(II) Optimism about a conception of responsibility.

(II) presupposes (I). But (I) does not imply (II). This leaves it open that Minimal Optimism is *part of* Strawson's view. Thus, the senses in which it, by itself, makes good on Strawson's Promise are truly ways in which *Strawson's approach* makes good on that promise. Strawson himself may also embrace an optimism about our particular conception of responsibility. This is compatible with, but not implied by, what we have been concerned to elucidate in this thesis. Formally, then, there is no tension in treating Minimal Optimism as (part of) Strawson's view, even if—which we'll consider next—Strawson is an optimist about our conception.

IV. Strawson's Optimism Reconsidered

Next, we are to consider Strawson's optimism about our actual conception of responsibility. In particular, we will consider how he in fact supports any such optimism, if it is any such optimism that he in fact supports.

Strawson is typically considered to be not only, in our terms, a Minimal Optimist but (also) an optimist about our particular conception. This is not unfounded. He is evidently *not* a sceptic about responsibility. And the libertarian idea of freedom he unequivocally dismisses as unintelligible (e.g., Strawson, 1998e, p. 170; 2008a, p. 27; Strawson et al., 2008, p. 93f.).²³⁹ Indeed, at least twice does he, if somewhat evasively, accept to be considered a “mere” or “straight compatibilist” (Strawson, 1995, p. 431; 1998e, p. 170).²⁴⁰ And at several places other than ‘Freedom and Resentment’ (Strawson, 1992a, ch. 10; 1995, pp. 430-432; 1998b, 1998e, 2011c), he explicitly argues that when we blame an agent, we take it that the agent could have acted otherwise, and that this in no way involves “an explicit or implicit denial of a thesis of determinism” (Strawson, 1995, p. 431). This does indeed sound like a straightforward compatibilism or optimism about responsibility (cf., De Mesel, 2022c, p. 5ff.).²⁴¹ By his own word, it was also this that he was concerned to show in ‘Freedom and Resentment’ (e.g., Strawson, 1998b, p. 261).²⁴²

However, the impression of optimism also owes, in some measure, to his quite dismissive attitude towards ‘philosophers’.²⁴³ To take one, most pertinent, example:

we can relax: the whole issue between determinists and libertarians is an irrelevance; and the fact that it has been so long and earnestly debated is but

²³⁹ It is to Kant’s credit, Strawson remarks, that Kant did not claim to understand it himself—“The most he claimed was that we could comprehend its incomprehensibility” (Pivčević, 1989, p. 7).

²⁴⁰ “If, in saying the above, I have dwindled into a mere compatibilist—in the company, say, of Hume—I am content with that” (Strawson, 1995, p. 431) and “So do I emerge as a straight compatibilist? If so, *ainsi soit-il*!” (Strawson, 1998e, p. 170).

²⁴¹ Others have noted (Alvarez, 2021, p. 188, n. 10; Watson, 2008, p. 119, n. 7) that it is not an altogether straightforward matter to declare Strawson a compatibilist. In light of his scepticism about the very intelligibility of the thesis of determinism and his rejection of the fundamental terms of debate between the (classic or instrumentalist) compatibilist and the libertarian, that label might be misleading. Perhaps this is just what his evasiveness regarding the term signals.

²⁴² For an account of how Strawson’s later writings on responsibility fit with his earlier (in particular, that is, ‘Freedom and Resentment’), see De Mesel (2022c, pp. 10-15).

²⁴³ This is something he shares not only with Wittgenstein (from whom he might have inherited some of this attitude), but also, of course, with Nietzsche. This is just to say that also the same ‘spirit’ may come in rather different guises.

one more illustration of the tendency of philosophers to raise a dust and then complain they cannot see. (Strawson, 1998e, p. 170)²⁴⁴

Also the failure, more generally, to get a clear view of morality, Strawson sees as owing to a character flaw of philosophers, viz.

the characteristically philosophical compulsion to seek and find the roots of the institution [of morality] in some single unitary source, be it reason, God, human emotions, social utility or some other philosophical darling. (Strawson, 1995, p. 433)²⁴⁵

Something of the same dismissive attitude is also present when he is asked to comment on “the commonest general criticism” of his work, viz., that to (merely) *describe* our conceptual scheme is conservative; in response, Strawson declares, “if it is conservative to resist essentially ineffective philosophical dreaming, [...] then I’m all for conservatism” (Magee & Strawson, 1971, p. 126). By Strawson’s lights, so it seems, the worries over free will, as some other philosophical worries, are primarily revealing of the hereditary defects of philosophers.²⁴⁶

Plausibly, however, these worries are not only that. Indeed, in his response to Bennett (1980), Strawson (1980, pp. 264-266) admits this. Bennett’s explanation of why we are so worried about free will, following Williams (1973), is in terms of the influence of Christianity, which, among other things, encourages the thought that “‘moral worth must be separated from any natural advantage whatsoever’” (Bennett, 1980, p. 26; B. Williams, 1973, p. 228).²⁴⁷ In his response, Strawson starts by conceding to Bennett both that a diagnosis of sceptical worries as stemming just from superficial conceptual muddles cannot be the full story and that the particular explanation that Bennett offers

²⁴⁴ Recall also that disdainful comment in ‘Freedom and Resentment’: “Sometimes he plugs this gap [between the applicability of some notion of libertarian freedom in particular cases and its moral consequences] with an intuition of fittingness—a pitiful intellectualist trinket for a philosopher to wear as a charm against the recognition of his own humanity” (2008a, p. 25).

²⁴⁵ I must say, given the issues here discussed, these words could have been Nietzsche’s.

²⁴⁶ So it seems, and so it is in part, but it should not give the wrong idea. Quips in short replies aside, ‘Freedom and Resentment’ (2008a) and ‘Freedom and Necessity’ (Strawson, 1992a?, ch. 10), at least, seem to be attempts not merely at *dissolving*—as showing of the problem as merely illusory would suggest—but rather at *solving* the problem. Strawson himself dates a change of mind on his part, from thinking that what we should do is dissolve philosophical problems, to thinking that what we should do is to solve philosophical problems, to the time when he was working on *Individuals* (see, Magee & Strawson, 1971, p. 118).

²⁴⁷ The explanation is omitted in “Accountability (II)” (2008).

is a convincing one. Strawson admits, that is, that the problem of free will is not simply an invention of philosophers, nor simply a matter of superficial conceptual confusion, but that it is, plausibly, a conceptual confusion induced or encouraged by our particular history. In conceding this, the discussion is clearly located at the level of our actual, socio-historically local conception of responsibility. For this reason, the exchange with Bennett is the most illuminating for our purposes. It is worth citing Strawson's reply at length:

He [Bennett] justly remarks that what, in 'Freedom and Resentment', I called the pessimist's response to the thesis of determinism cannot be explained as solely the consequence of *superficial* conceptual muddles about the requirements of freedom. He produces a convincing deeper explanation [...], in part historical, in part more generally natural, of why accountability, and the appropriateness of reactive feelings, should be thought, not only by libertarians but also by sceptics like Ayer, to require some ultimate, and ultimately unintelligible, kind of 'freedom'. He does not dwell as much as he might on the oddity of this view. [...] One is, indeed, inclined to say that the incoherence of the alleged condition infects the thought that the appropriateness or rationality of the reactive attitude requires its fulfilment. Bennett himself is prepared, or half-prepared, to allow this thought some place, though not a dominant place, in 'our ordinary concept of accountability'. But there is a quite general ambiguity in the notion of 'our ordinary concept' of whatever it may be. Should the lineaments of such a concept be drawn exclusively from its use, from our ordinary *practice*, or should we add the reflective accretions, however confused, which, naturally or historically, gather around it? The distinction is hardly clear-cut; but where it can be made, I prefer the first alternative.

What I was above all concerned to stress [in 'Freedom and Resentment'] was that our proneness to reactive attitudes is a natural fact, woven into the fabric of our lives, given with the fact of human society as we know it, neither calling for nor permitting a general 'rational' justification. We can see where the limits of our proneness to these attitudes tend to fall, we can understand why they tend to fall where they do and we can find room for the idea of criticism, of appropriateness and inappropriateness, in particular cases. That is all; and that is enough. (Strawson, 1980, p. 264f., original emphasis)

There is a lot in this passage. First, we have the already mentioned concessions, that there is something right in Bennett's (and Williams') diagnosis, culminating in the suspicion that "the incoherence of the alleged condition [i.e., of ultimate freedom] *infects* the thought that the appropriateness or rationality of the reactive attitude requires its fulfilment" (emphasis added). Second, we have a distinction, between the ordinary practice and the (at least partly) socio-historically particular reflective thoughts we might have in connection to that

practice. Last, we get the core of Strawson's position in 'Freedom and Resentment', that we are naturally committed to the concept of responsibility.

The concessions are important for us only because they set the stage, or the level, of the discussion. The distinction Strawson introduces we will attend to in a bit. First, we should observe another quite general ambiguity in the notion of 'our ordinary concept', one we have stressed before. We need to distinguish between an *inclusive* and a *contrastive sense* of 'our' or 'us'. The last paragraph of the quote represents a *slide* from the one sense to the other. It appeals to the idea, as we've explicated it, that we are inescapably committed to *the concept* of responsibility, but not to any particular *conception* of responsibility. In different terms, what is "given with the fact of human society", we've seen, is our concept of responsibility where 'our' is understood *in an inclusive sense*; what does not follow, however, from "the fact of human society" is *our* concept of responsibility where 'our' is understood *in a contrastive sense*—i.e., our conception of responsibility in contrast to the conception of some other human society. If the concessions really do place the discussion at the level of *our particular conception* of responsibility, then there is, with the last paragraph of the quote above, a *slide* in Strawson's argument to a different level, viz. to that of the concept of responsibility. What does this say about Strawson's answer? And what does it say about his optimism?

On one way of reading Strawson's response, he appears as *genealogically insensitive*. He seems insensitive to the socio-historical particularity of our conception of responsibility. With 'the slide' it seems like Strawson mistakenly takes what is in fact only concept-optimism (i.e., his Minimal Optimism) to ground a conception-optimism.²⁴⁸ On this reading, the distinction he appeals to in identifying 'our ordinary concept'—between the practice and the socio-historically peculiar reflections we are prone to in connection to this practice—appears as an attempt to purge our conception from what seemed socio-historically contingent about it. That is, Strawson seems to be saying that our actual practice is not so socio-historically peculiar as we might have thought. It is quite ordinary, in that sense—"This is what human beings everywhere and always have been doing, nothing strange about it". The distinction serves to identify, in the terms we are using, *our conception* with *the concept*.

This reading fits with the fact that Strawson's primary concern is with the fundamental core of human thinking. But it could also have us wondering whether Strawson is in fact to be considered an optimist about our conception

²⁴⁸ It presents him only as *somewhat* insensitive because this reading does take the distinction between the concept and any conception to be central, and this does presuppose *some* sensitivity to the socio-historical particularity of different conceptions.

of responsibility. For to think that *our particular conception* is identical with (or even near-identical with) *the concept* is to show a serious lack of historical sense. The passages we have considered in elucidating his Minimal Optimism suggest that Strawson isn't lacking in historical sense.

A more charitable reading is that Strawson does not in fact take his Minimal Optimism to adequately ground an optimism about our conception of responsibility. So understood, 'the slide' is rather *a retreat*—a retreat to a higher level of generality. On this reading, the conditions of appropriateness for the reactive attitudes that he does outline, in 'Freedom and Resentment' and elsewhere, are best left (as they are) very schematic. They reflect only *the contours* of our conception—i.e., that which follows simply from the shape of *the concept*. Within these contours, different conceptions may fill in the conditions differently, but it is the fact that there are these contours to be filled in that is important. "That is all; and that is enough". Strawson's letter to Russell might be taken to suggest this reading.²⁴⁹ As does some remarks in his response to Bennett, following the passage quoted—for example, "I'm driven to think that he [Bennett] construes some of the phrases I used [in 'Freedom and Resentment'] as having a greater definiteness than I intended them to have or, perhaps, than the case admits of" (Strawson, 1980, p. 266).

On this more charitable reading, however, we drop the assumption that Strawson is an optimist about our conception of responsibility. He is *only* a Minimal Optimist. He may nevertheless seem rather genealogically insensitive. That is, for this perspective, he seems insensitive to how our concerns about responsibility are rooted in a socio-historically particular conception, thinking that the worries are adequately dealt with as long as Minimal Optimism is granted. This is a less serious charge than the charge of lacking historical sense, and more plausible with respect to Strawson. On this reading, we might preserve a sense in which he is an optimist about our particular conception of responsibility: he is *negatively* optimistic, in the sense that he expresses *no discomfort* with our actual conception of responsibility.

Another possible reading is that 'the slide'—from the level of our conception to the level of the concept—serves to remind us of why we have any conception of responsibility and thus effect a more realistic understanding of our own conception. It is not, then, a flat-footed conflation, nor merely a retreat to a more comfortable level of analysis. While the idea behind Minimal

²⁴⁹ "The distinction you [Russell] draw between the 'rationalistic' and the naturalistic strategies is too sharp. It is, rather, just another natural fact that 'excusing conditions' referred to under the former head are humanly accepted as such and that their acceptance as such is generally approved" (Appendix).

Optimism does not *ground* an optimism about our conception, it serves an *analytical* purpose in pursuit of such optimism. By pointing to the natural fact and, in particular, *why* it is a natural fact about us that we have the concept of responsibility—its *functional core*—we are brought, in trying to understand our own conception, to a more *concrete* and *realistic* understanding of our concepts.

To see this, consider, first, the distinction which Strawson invokes, between our ordinary *practice* and our *reflective thoughts*. As he admits, the distinction is not clear-cut. On the assumption that we are concerned with a particular *conception* of responsibility, the question seems to be whether we should understand a given conception, including what is socio-historically peculiar about it, in terms of a socio-historically local *practice*, or if we should add to our understanding of the conception also the reflective thoughts these particular concept-users (of that time and place) are prone to have. The distinction is not clear-cut because reflection is also something we *do*. It is not plausible that these reflective thoughts about responsibility (however socio-historically peculiar they may be) would be completely epiphenomenal to (the rest of) our practice of responsibility.

The *point* of the distinction, however, seems to be that, while “the alleged condition [i.e., of ultimate freedom] infects the thought that the appropriateness or rationality of the reactive attitudes requires its fulfilment” (Strawson, 1980, p. 265), this thought does not, in fact, play a major role for us in practice. In trying to achieve conceptual self-understanding, we are prone to look too narrowly, taking these somewhat peculiar reflective thoughts to express necessary requirements of our conception of responsibility, when in fact they are not particularly important for how we actually act.

Another distinction, already implicit in the first, seems better suited for this point: between, on the one hand, some *special*, possibly more metaphysically demanding conception and, on the other hand, our *ordinary* conception. Rather than saying that (many of) the reflective thoughts we are prone to are epiphenomenal, we should say that the conception we actually use most of the time—our *ordinary* conception—is not as these reflections make it out to be, or indeed require of it that it is. The lineaments of our *ordinary* conception should be drawn from our *ordinary* practice. The particular and indeed peculiar reflective thoughts at issue are not central for our ordinary practice, nor for our ordinary conception. We may still say that these reflections express a different—a special, perhaps more metaphysically demanding—conception of responsibility; one which we also have. The pessimist’s mistake is to think that it is this special conception that needs to be vindicated. (Consider: “He [Bennett] does not dwell as much as he might on the oddity of this [Christianity

influenced] view” (Strawson, 1980, p. 264f.).) So understood, Strawson’s optimistic attitude is based on the thought that our ordinary conception is *not* this special conception; that it is the ordinary conception, the one which we (most often) employ in practice, that is important; and that this conception serves us quite well in the relevant respects.

The point of ‘the slide’—from our conception to the concept—is to remind us of the *functional core* of the conception. That is, in his response to Bennett, immediately after pointing to the primacy of our ordinary practice, Strawson reminds us of why it is that we have *any* conception of responsibility. He points to a natural fact about us—that we have *some* conception of responsibility—and how, in seeing it as such, “we can understand why [the limits of our proneness to these attitudes] tend to fall where they do”. As we’ve explicated it, our framework commitment to responsibility is intelligible as practically necessary given certain basic needs present in any human society. This practical role—the *functional core* of any conception of responsibility—is what Strawson reminds us of. Our own conception is an elaboration of, or development from, that core. By reminding us of what our particular conception fundamentally does for us, Strawson seeks to bring us, in Wittgenstein’s (2009a, § 107) phrase, ‘back to the rough ground’.

A point in favour of understanding Strawson as doing just this in his response to Bennett is that he has explicitly made this move before. In the last chapter, we saw that Strawson, in ‘Social Morality and Individual Ideal’ (2008d), constructs ‘a minimal conception of morality’, on which morality is represented in the first instance as “a kind of public convenience” (2008d, p. 34). He is clear that this minimal conception is not supposed to adequately describe morality as we know it; there is no illicit equation (or mistaken conflation) here. Strawson’s claim is “only that it is a useful analytical idea” (ibid.). He presents this schematic account of ‘morality’ in terms of its functional core. In terms familiar from the method of pragmatic genealogy, by asking why human beings who did not already have a moral system would need such a system, we’re presented with an account of why a *generic* version of the institution of morality would develop given very basic human needs and situations. This generic institution is not an adequate representation of our, or any other, *actual* institution of morality, but it does elucidate why it is practically necessary for us, given these basic needs and situations, that we have *some* such institution.

A virtue of constructing this minimal, schematic account—the generic version of morality—Strawson claims, is that it helps us understand our actual concepts “in a concrete and realistic way” (2008d, p. 37). He illustrates this by reference to Elisabeth Anscombe’s argument in ‘Modern Moral Philosophy’

(1958) that our concept of obligation is unintelligible without a Christian outlook. (Of course, Anscombe's diagnosis shares much with Williams', which Bennett appeals to.) In light of the minimal construction of morality, Strawson's response to Anscombe is that our *ordinary* concept is not like that.

the present approach [...] makes it relatively easy to understand such notions as those of conscientiousness, duty and obligation in a concrete and realistic way. These notions have been treated almost entirely abstractly in moral philosophy in the recent past, with the result that they have come to some of our contemporaries [(cf., Anscombe, 1958)] to seem to be meaningless survivals of discarded ideas about the government of the universe. But as most ordinarily employed I do not think they are all that. There is nothing in the least mysterious or metaphysical in the fact that duties and obligations go with offices, positions and relationships to others. [...] [W]hen we call someone conscientious or say that he has a strong sense of his obligations or of duty, we do not *ordinarily* mean that he is haunted by the ghost of the idea of supernatural ordinances; we mean rather such things as this, that he can be counted on for sustained effort to do what is required of him in definite capacities, to fulfil the demand made on him as student or teacher or parent or soldier or whatever he may be." (Strawson, 2008d, p. 37, emphasis added)

By apprehending the functional core of any moral system, we are helped to a more concrete and realistic understanding of our actual moral concepts.²⁵⁰ Human beings need to be able to rely on other people in pursuing their own ends and, partly just for this reason, we have a need to differentiate between reliable and unreliable people. It might be that we (without our particular ethical outlook) employ the concept of moral obligation in several different ways, but *as most ordinarily employed*, Strawson stresses, it remains connected to that basic function.

In response to Anscombe, the suggestion is, as in his response to Bennett, Strawson thinks we achieve a more concrete and realistic understanding of our own conception if we attend to the fact that we necessarily have *some* conception of that sort and *why* it is that we necessarily have that. Our ordinary conception of responsibility is a development of, or elaboration from, the functional core of the concept of responsibility; if we see our conception in light of why human beings (anywhere and everywhere; i.e., already in the State-of-Nature) would need to have *some* such conception, rather than trying to understand our conception in terms of the 'reflective accretions' that have

²⁵⁰ See also, "This would be one way [...] of using the minimal idea of morality to try to get clearer about the ordinary idea" (Strawson, 2008d, p. 34).

gathered around the concept, we may better recognise what the conception actually expressed in our ordinary practice is like.

On this reading, Strawson *is* an optimist about our conception of responsibility. Certain ‘reflective accretions’ may have gathered around our notion of responsibility, but they are not important for the ordinary way in which we employ that concept. By being insufficiently attentive to what, at bottom, our conception of responsibility does for us, we are prone to make it out as something very special, which we then struggle to square with our naturalistic conception of the world. The idea behind Strawson’s Minimal Optimism—of the indispensable functional role of the concept of responsibility—thus serves a significant *analytical* role in supporting an optimism about our conception of responsibility. It serves to provide us with a more concrete and realistic understanding of our ordinary conception of responsibility.

Also on this reading, we might object that Strawson’s optimism is unwarranted—if this is really all, it is certainly not enough. In this light, Strawson still appears somewhat genealogically insensitive. But he is at least not simply evasive; there is a positive, and indeed critical, aspect to invoking our natural commitment to responsibility. We may count this as a further anti-sceptical potential of Minimal Optimism.

V. Strawson’s Realism and the Limits of Description

There is more of a critical aspect to the last reading than Strawson himself seems to admit that his approach has. By vindicating only ‘the ordinary conception’, the approach seems, in Bennett’s (2008, p. 55) words, “more revisionary—or rather, excisionary”—than Strawson lets on. This seems right. The rather peculiar conceptions are plausibly at least part of our actual ethical outlook, given that this outlook in fact is the product of a certain history. Given that this is the actual state of our ethical outlook, to only describe settles almost nothing. In doing moral philosophy, we face the limits of description.

While we have not strayed from the text, *the spirit* of this reading might seem not to be Strawson’s own. To some extent, this is admittedly so. But this suspicion should not be exaggerated. Not even the spirit of this reading might in fact be that far from Strawson’s own—at least if we attend to what he says rather than what he says he says.

Consider, once more, what he actually says in that interview by which we introduced the Minimal Optimist idea:

Insofar as freedom is linked to responsibility, then indeed we are naturally committed to the belief in freedom of action [...] this doesn't involve some metaphysical, and indeed unintelligible, libertarian conception of freedom ... but it does involve all that we actually need, or can hope for, in the way of responsibility and freedom. (Strawson et al., 2008, p. 93f.)

According to Strawson, it is not only the case that we have only what we *need*, as opposed to what we might have *wanted*; the claim is that we have only what we *actually* need, as opposed to what we might have *thought* we needed; we have what we *can* hope for, which does not mean that we have what we might actually have hoped for.

The same sentiment is expressed in another interview, around the same time:

We make decisions and choices in light of our desires and beliefs, and if nothing prevents our acting in accordance with those choices and decisions, that is *all the freedom it makes sense to want*. (Pivčević, 1989, p. 7, emphasis added)

That we have all the freedom *it makes sense to want* does not mean that we have all the freedom that we actually (however confusedly) want.

These are the words of a *realist*, in the colloquial sense of that term—i.e., the sense of that terms which stands in self-congratulatory contrast to 'optimism'. This realistic attitude complicates the picture of Strawson as a complacent compatibilist. It does so for precisely the reason that it complicates the picture of him as merely describing. There indeed seems to be a more critical edge to Strawson's actual views on responsibility than usually recognised; indeed, than he himself is prone to admit.

Strawson's aim "to exhibit [our conceptual scheme ...] as a coherent whole whose parts are mutually supportive and mutually dependent, interlocking in an intelligible way" might be valid at the level at which he most often pursues it, i.e., with respect to "the major structural elements or features of our conceptual scheme" (Strawson, 2008c, p. 18). But it is not plausible that there is this level of coherence in ethics, when this is concerned with the conceptions of an actual ethical outlook. Actually, that seems to be Strawson's own view:

it is as wholly futile to think that we could, without destroying their character, systematize these truths [i.e., the pictures of man that capture the ethical imagination] into one coherent body of truth as it is to suppose that we could, without destroying their character, form a coherent composite image from these images. This may be expressed by saying that the region of the ethical is the region where there are truths but no truth; or, in other words, that the injunction

to see life steadily *and* see it whole is absurd, for one cannot do both. (Strawson, 2008d, p. 32, original emphasis)²⁵¹

With respect to the ethical, there is no ‘coherent whole’ to be described. If it can be coherently described, it will only express *a* truth, not *the* truth. At least, there is not enough coherence for there to be much to be *merely described*. To seek to vindicate, what Strawson considers, our *ordinary* conception and to free us from, what he considers, some peculiar and rather removed conception, that is to take a stance—to express *a* truth—and not, or not merely, to describe.²⁵² Herein lies the truth, I think, in Bennett’s suspicion that Strawson’s approach is in fact “more revisionary—or, rather, excisionary”—than Strawson’s own description of what he is doing might lead one to think.²⁵³

VI. Conclusion

This chapter has ventured both a bit and quite far beyond Strawson’s Minimal Optimism. In two respects, the Minimal Optimist idea has been presented as more critical than we might have thought. For one thing, while Minimal Optimism is unapologetically inadequate for defending a particular ethical outlook (such as ours) and, which is the other side of that coin, for rejecting scepticism about a particular conception of responsibility (such as ours), there is nevertheless a critical edge to it. Minimal Optimism stands in recognition of the fact that our actual conceptions of responsibility, blame, and freedom have

²⁵¹ Cf. “It seems too readily assumed that if we can only discover the true meanings of each of a cluster of key terms, usually historic terms, that we use in some particular field (as, for example, ‘right’, ‘good’ and the rest in morals), then it must without question transpire that each will fit into place in some single, interlocking, consistent, conceptual scheme. Not only is there no reason to assume this, but all historical probability is against it, especially in the case of a language derived from such various civilisations as ours is. We may cheerfully use, and with weight, terms which are not so much head-on incompatible as simply disparate, which just don’t fit in or even on. Just as we cheerfully subscribe to, or have the grace to be torn between, simply disparate ideals—why *must* there be a conceivable amalgam, the Good Life for Man?” (Austin, 1956, p. 29, n. 16).

²⁵² While giving a lecture in Communist-era Yugoslavia, Strawson was accused by a member of the audience of having a bourgeois outlook; his response: “But I am bourgeois—an elitist liberal bourgeois” (Strawson, 1998a, p. 16). See O’Grady (Krishnan, 2023, p. 311; O’Grady, 2006; Strawson, 1998a).

²⁵³ This is not how Bennett understands Strawson. He thinks “Strawson means to claim only to have provided for every *coherent* element in ‘what we mean’, so that what is offered is not a fully conservative theory but rather a maximal salvage” (1980, p. 25, original emphasis).

a history. In recognising this, Strawson's Promise, in so far as this is understood as that of taking us beyond the traditional debate, has been further substantiated: scepticism about our particular conception should not land us in the nihilistic conclusion that no conception of responsibility will be liveable.

Venturing further beyond Minimal Optimist, we considered Strawson's optimism about our particular conception of responsibility; how he in fact supports any such optimism, if it is any such optimism that he in fact supports. It might be that Strawson is only a Minimal Optimist or at most an optimist about our conception of responsibility in a negative sense, in that he rests assured, somewhat complacently, in the belief that Minimal Optimism is enough. But it does seem like Strawson is in fact an optimist also about our conception of responsibility, and possibly a more critical one than generally assumed. The idea behind Strawson's Minimal Optimism was shown to serve an analytical purpose in helping us see the practical role of our concepts, awarding us with a more concrete and realistic understanding of them, salvaging our ordinary conception from the threat of scepticism by separating it from what seem like loitering ideas inherited from an enchanted world.

Admittedly, the way Strawson's approach has been presented in this chapter differs somewhat from the spirit in which he himself presented it. Towards the end, said something to suggest that, although Strawson presented his view differently and, in particular, took himself to be engaged only in description, and not revision, of our conceptual scheme, we have reason to be critical of this presentation, given what Strawson actually said, in contrast to what he said that he was saying.

8. Conclusion

It has been said that the best conceptual scheme, the best system of ideas, is the one that gets us around best. The question is: in what milieu?

P. F. Strawson, 'Two Conceptions of Philosophy'

The reason 'Freedom and Resentment' has been so influential is that it evinces the sense that we may move beyond the traditional worries about free will, recognizing that the truth of determinism is irrelevant for our understanding of ourselves as responsible beings. This is the promise of Strawson's approach. This thesis has explored whether and how we can make good on it.

In the first part of this thesis, we considered three different ways in which we might have thought that something like Strawson's Promise could be made good on: a transcendental argument, the Statistical Reading, and the Reversal Move.

In chapter 2, we saw that the transcendental argument which Justin Coates suggests that Strawson advances in 'Freedom and Resentment' cannot make good on Strawson's Promise, at least not as *Strawson's* Promise. The transcendental argument he ascribes to Strawson is *the wrong kind* of transcendental argument (a *deductive* rather than an *elenctic* argument) and, more fundamentally, the anti-sceptical approach he ascribes to Strawson is *the wrong kind* of anti-sceptical approach (a *direct* rather than an *indirect* approach).

In chapter 3, we saw that the motivation for Hieronymi's novel reading of 'Freedom and Resentment', the Statistical Reading, was confused. If we attempt to solve the Incoherence Puzzle per Hieronymi's suggestion, we ensnare ourselves in further interpretative puzzles. The Statistical Reading proves not to be a coherent interpretation of 'Freedom and Resentment'. It is therefore unpromising as a way of trying to make good on Strawson's Promise.

In chapter 4, we considered two accounts of the Reversal Move—the Fitting Response-Dependence Thesis and the Meaning-Based Account. While the Meaning-Based Account has a good claim to being Strawson's view, it does

not make good on Strawson's Promise. And the Fitting Response-Dependence Thesis, whether it is adequately anchored in Strawson's approach or not, leaves the important work left to be done. What is needed, I concluded, is a naturalistic account that shows that determinism is not relevant for responsibility.

In the second part of the thesis, we explored the Minimal Optimist Reading of Strawson's approach. At the centre of this approach is the Inescapability Claim, properly construed: that we are naturally and inescapably committed to the concept of responsibility, even if no particular conception of responsibility is in that sense inescapable. The concept is *conceptually necessary* in the sense that it is essential if we are to have anything that we can make intelligible to ourselves as a system of human relationships. Unlike other framework commitments, however, our commitment to the concept of responsibility is *not* "essential to any conception of experience which we can make truly intelligible to ourselves" (Strawson, 2019, p. xix, emphasis added). We do have recourse to a kind of experience which does not make essential use of this concept—i.e., the objective attitude. For this reason, with respect to the anti-sceptical force of the Inescapability Claim, pressure is put on a further sense in which the concept of responsibility is inescapable. The concept of responsibility is *practically necessary* in the sense that it is indispensable for beings like us simply in virtue of very basic needs and situations; it is given, already, "with the fact of human society" (Strawson, 2008a, p. 25). By supplementing Strawson's approach with the method of pragmatic genealogy—that is, with a naturalistic explanation that has a good claim to being of the kind that Strawson recognised as necessary for philosophical understanding—we can support the Minimal Optimist idea by rendering the inescapability of the concept of responsibility intelligible as indispensable for beings like us. The pragmatic genealogy vindicates us having *some* conception of responsibility, even if it does not—simply at the level at which we considered it—vindicate any particular conception of responsibility.

The most basic sense in which Minimal Optimism may be said to make good on Strawson's Promise concerns the attitude that Strawson evinces. As Anil Gomes (2019) notes, 'Freedom and Resentment' embodies "one of the central themes of Strawson's work: a relaxed sympathy for our ordinary ways of thinking about ourselves and our role in the world". On the reading here proposed, with respect to responsibility, this relaxed attitude is fundamentally grounded in a Minimal Optimism—that is, in the recognition of the fact that we are naturally and inescapably committed to having a conception of responsibility.

Another way of putting this is that, the sense in which Minimal Optimism makes good on Strawson's Promise is that it displaces the question of *whether* we are responsible. We are responsible. The concept of responsibility is an inescapable feature of any human conceptual scheme. We cannot do without it. So, there is no point in asking whether we are responsible, *as if* we could actually go about conceiving of ourselves in any other way. In this sense, Minimal Optimism brings us beyond the traditional debate. It displaces scepticism about responsibility. Or, more precisely, it displaces scepticism about *the concept* of responsibility—what we called *nihilism* about responsibility.

Because Minimal Optimism, while a form of optimism, nevertheless allows for the possibility of scepticism about any particular conception of responsibility, it helps us gain critical distance to our own conception; it helps us to achieve an historical sense. It averts the privileging of our own conception as “the true” conception. In so doing, it also averts the illicit inference from scepticism about our conception to nihilism about responsibility. Scepticism about our conception does not mean that we are not responsible, but rather means that we are landed with the serious task of finding another conception by which to live.

Not just anything we might wish to call ‘responsibility’ will do, however. Importantly, already the concept of responsibility must be backward-looking. We have seen that for the practice of blame to be functional, a precondition is that the reasons for why the concept is applied in any particular case must be autonomous, i.e. not conditional on the function of the practice. Already the generic version of the practice of blame—that which follows simply from within our State-of-Nature, on the assumption only of maximally generic needs—must be, if it is to be functional, autonomous. Any conception of responsibility that is to sustain the practice of blame must thus have a backward-looking character. This is important for reasons having to do with the contemporary discussion about free will, because the sceptic's favourite alternative to our actual conception of responsibility is a conception of responsibility that is purely forward-looking or instrumental. In light of the pragmatic genealogy of blame, this appears, in Strawson's words, as “essentially ineffective philosophical dreaming” (Magee & Strawson, 1971, p. 126).

We might however be concerned about Strawson's relaxed attitude being a bit *too* relaxed. Is his optimism not only that of Minimal Optimism, but more substantial—if so, what ground does he have for such optimism? In the last chapter, we saw how Strawson might rest some of his more substantial optimism on the idea behind Minimal Optimism. In light of *why* we necessarily

have a conception of responsibility, we grasp what our conception of responsibility is most fundamentally doing for us. Whether we can, for that reason, relax, I'm personally not sure.

Strawson's *Scepticism and Naturalism* bears the following epigraph, from the 18th century historian Edward Gibbon:

The satirist may laugh, the philosopher may preach; but reason herself will respect the prejudices and habits, which have been consecrated by the experience of mankind.²⁵⁴

Can we trust “the prejudices and habits” of mankind? In asking this question, I'm not giving voice to any general suspicion about our form of life having been “consecrated by the experience of mankind”. The question concerns, rather, the issue of whether we can trust that *these* practices—that is, *their* prejudices and habits. Since the formation has taken place *in the past*, will these practices and ideas actually serve *us*, both now and going forward? In what milieu do we need to, and do we want to, get around in? What is the best conceptual scheme *for us*?

A famous passage from J. L. Austin's ‘A Plea for Excuses’ reads:

our common stock of words embodies all the distinctions men have found worth drawing, and the connexions they have found worth marking, in the lifetimes of many generations: these surely are likely to be more numerous, more sound, since they have stood up to the long test of the survival of the fittest, and more subtle, at least in all ordinary and reasonably practical matters, than any that you or I are likely to think up in our armchairs of an afternoon (1956, p. 8).

The question is whether we can simply relax and be comfortable with the stock of words that generations past have found helpful, knowing very well that we do *not* share their concerns in some (and some very important) respects.

What has been said for Strawson's relaxed approach is, in this respect, not enough. It is not enough for adequately grounding reasonable confidence in our actual conception of responsibility. Minimal Optimism answers to the suspicion just raised—a suspicion about *the Wisdom of the Ages*—by pointing out that, in just the respects in which it is optimistic, in just these respects we can really trust in “prejudices and habits [...] consecrated by the experience of mankind”. The concept of responsibility, in its fundamental form, is inescapable simply in virtue of maximally generic needs and situations. On a properly delimited Inescapability Claim, Strawson's relaxed attitude is

²⁵⁴ From Gibbon (2004) *Memoirs of My Life and Writing*.

grounded in a realistic understanding of our common form of life. This takes us beyond the traditional debate, I've suggested, but it is not, by itself, enough. In so far as Strawson thought it was—which, at some points, he seems to have thought—then he appears, in this light, as genealogically insensitive; that is, for one thing, as insensitive to how our concerns about responsibility and freedom are rooted in our particular conception of responsibility and, second, as insensitive to the fact that, while there is a common core, the differences between conceptions of responsibility are significant.

On the Minimal Optimist picture, it is not a question *whether* we are responsible. We are. In light of this, however, the central question becomes: What does it *mean* to be responsible? Or, in other words, what conception of responsibility do we, and should we, have? On the most charitable reading of Strawson's approach, while we may be helped in gaining a more concrete and realistic understanding of our conception, he does not adequately answer this question. That is up to us. It is part of Strawson's legacy that this is just the question to which philosophers have turned their attention in recent years. This thesis has sought to vindicate this reorientation as a consequence of Strawson's approach, thus showing that the promise it suggested was not illusory.

Appendix: Letter to Paul Russell, from P. F. Strawson, 1992

Typescript

Magdalen College
Oxford OX 14AU
Telephone (0865) 276000

Dear Professor Russell,

Thank you for sending me your admirable article.* I will allow myself just two comments.

(1) I do not want to rule out the idea of universal incapacitation as incoherent or logically impossible. But the thoroughgoing Pessimist must hold that the truth of determinism would have universal moral incapacitation as a consequence. I do not believe that and neither, I think, do you.

(2) The distinction you draw between the 'rationalistic' and the naturalistic strategies is too sharp. It is, rather, just another natural fact that 'excusing conditions' referred to under the former head are humanly accepted as such and that their acceptance as such is generally approved.

[page break]

Finally, though it is true that I make no explicit attempt to give a 'positive account' of 'moral capacity', I think it is fair to say that such an account is more or less implicit in the piece as a whole.

Forgive the slightness of these comments; and let me add that I greatly appreciated the care and clarity of your admirably reasoned paper.

Your sincerely,
P. F. Strawson

* Russell (2017d). Originally published as Russell, Paul (1992) "Strawson's Way of Naturalizing Responsibility", *Ethics*, 102, pp. 287-302.

MAGDALEN COLLEGE
OXFORD OX1 4AU
Telephone (0865) 276000

August 9th 1992

Dear Professor Russell,

Thank you for sending me your
admirable article. I will allow myself
just two comments.

(1) I do not want to rule out the idea
of ~~universal~~ universal incapacitation as
incoherent or logically impossible. But the
thoroughgoing Pessimist must hold that the
truth of determinism would have universal
moral incapacitation as a consequence.
I do not believe this and neither, I think,
do you.

(2) The distinction you draw between the
'rationalistic' and the naturalistic strategies
is too sharp. It is, rather, just another
natural fact that the 'excusing conditions'
referred to under the former head are
humanly accepted as such and that their
acceptance as such is generally approved.

P.T.O.

Finally, though it's true that I make no explicit attempt to give a 'positive' account of 'moral capacity', I think it is fair to say that such an account is more or less implicit in the piece as a whole.

Forgive the slightness of these comments; and let me add that I greatly appreciated the care and clarity of your admirably reasoned paper.

Yours sincerely,

P.F. Strawson

References

- Allais, L. (2014). Freedom and forgiveness. *Oxford studies in agency and responsibility*, 2, 33-61.
- Allais, L. (2024). Navigating 'Freedom and Resentment'. In S. Heyndels, A. Bengtson, & B. De Mesel (Eds.), *P. F. Strawson and his Philosophical Legacy* (pp. 260-279). Oxford: Oxford University Press.
- Alvarez, M. (2021). P. F. Strawson, Moral Theories and 'The Problem of Blame': 'Freedom and Resentment' Revisited. *Aristotelian Society Supplementary Volume*, 95(1), 183-203.
- Annual Report, 1960-1. (1961). *Proceeding of the British Academy*, 47.
- Anscombe, G. E. M. (1958). Modern Moral Philosophy. *Philosophy*, 33(124), 1 - 19.
- Aristotle. *Metaphysics*. In.
- Arneson, R. (2003). The Smart theory of moral responsibility and desert. In S. Olsaretti (Ed.), *Desert and Justice*: Clarendon Press.
- Arpaly, N. (2006). *Merit, Meaning, and Human Bondage*: Princeton University Press.
- Austin, J. L. (1950). Truth. *Proceedings of the Aristotelian Society*, 24, 111-129.
- Austin, J. L. (1956). I.—A Plea for Excuses: The Presidential Address. *Proceedings of the Aristotelian Society*, 57(1), 1-30. doi:10.1093/aristotelian/57.1.1
- Baker, G. P., & Hacker, P. M. S. (2005). *Wittgenstein: Understanding and Meaning*: Blackwell.
- Baker, G. P., & Hacker, P. M. S. (2009). *Wittgenstein: Rules, grammar and necessity*: Wiley-Blackwell.
- Bakhurst, D. (2017). The Spirit of Pragmatism in the Quads of Oxford. In C. Misak & H. Price (Eds.), *The Practical Turn: Pragmatism in Britian in the Long Twentieth Century*: Oxford University Press.
- Baldwin, T. (Ed.) (1990). *G. E. Moore: Selected Writings*: Routledge.
- Bangu, S. (2021). Hard and Blind: On Wittgenstein's Genealogical View of Logical Necessity. *Philosophy and Phenomenological Research*, 102(2), 439-458.
- Bardon, A. (2005). Performative transcendental arguments. *Philosophia*, 33(1-4), 69-95.
- Bardon, A. (no date). Transcendental Arguments. *Internet Encyclopedia of Philosophy*. Retrieved from <https://iep.utm.edu/trans-ar/>

- Barrett, J. (2020). Optimism about Moral Responsibility. *Philosophers' Imprint*, 20(33), 1-17.
- Beglin, D. (2018). Responsibility, libertarians, and the “facts as we know them”: A concern-based construal of Strawson’s reversal. *Ethics*, 128(3), 612-625.
- Beglin, D. (2020). Two Strawsonian strategies for accounting for morally responsible agency. *Philosophical Studies*, 177(8), 2341-2364.
- Bengtson, A. (2019). Responsibility, Reactive Attitudes and Very General Facts of Human Nature. *Philosophical Investigations*, 42(3), 281-304.
- Bengtson, A. (2020). *The Concept of Moral Responsibility and Some Very General Facts of Nature: A Wittgensteinian Approach to Strawson’s ‘Freedom and Resentment’*. (PhD). KU Leuven, Leuven, Belgium.
- Bennett, J. (1966). *Kant’s Analytic*. Cambridge University Press.
- Bennett, J. (1980). Accountability. In Z. van Straaten (Ed.), *Philosophical Subjects* (pp. 14-47): Clarendon Press.
- Bennett, J. (2008). Accountability (II). In M. McKenna & P. Russell (Eds.), *Free Will and Reactive Attitudes*: Routledge.
- Berlin, I. (2009). *Enlightening: Letters 1946-1960* (H. Hardy & J. Holmes Eds.). London: Chatto & Windus.
- Berlin, I. (2017). *Affirming: Letters 1975-1997* (H. Hardy & M. Pottle Eds. Epub ed.). London: Pimlico.
- Blackburn, S. (1998). Relativization and Truth. In L. E. Hahn (Ed.), *The Philosophy of P. F. Strawson* (pp. 151-167): Open Court.
- Brentano, F. (1889/2009). *The Origin of Our Knowledge of Right and Wrong*. London: Routledge.
- Brink, D. O., & Nelkin, D. K. (2013). Fairness and the Architecture of Responsibility. In D. Shoemaker (Ed.), *Oxford studies in agency and responsibility* (Vol. 1): Oxford University Press.
- Broad, C. D. (1930). *Five Types of Ehtical Theory*. London: Routledgce & Kegan Paul.
- Broad, C. D. (1978). *Kant: An Introduction*. Cambridge: Cambridge University Press.
- Brown, C. (2006). *Peter Strawson*. Stocksfield: Acumen.
- Callanan, J. J. (2011). Making Sense of Doubt: Strawson’s Anti-Scepticism. *Theoria*, 77(3), 261-278.
- Campbell, J. (2017). P. F. Strawson’s Free Will Naturalism. *International Journal for the Study of Skepticism*, 7(1), 26-52.
- Carlsson, A. B. (forthcoming). Responsibility and the Emotions. In M. Kiener (Ed.), *The Routledge Handbook of Philosophy of Responsibility*: Routledge.
- Caruso, G. (2021). Skepticism About Moral Responsibility. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition ed.).
- Cassam, Q. (1997). *Self and World* (e-book ed.): Oxford University Press.

- Cassam, Q. (1999). Self-Directed Transcendental Arguments. In R. Stern (Ed.), *Transcendental Arguments: Problems and Prospects*. Oxford: Oxford University Press.
- Cassam, Q. (2008). Foreword. In *Scepticism and Naturalism: Some Varieties* (pp. vii-xviii): Routledge.
- Cassam, Q. (2016). Knowledge and its Objects: Revisiting the Bounds of Sense. *European Journal of Philosophy*, 24(4), 907-919.
- Chattopadhyaya, D. P. (1995). Categorical Frameworks. In P. K. Sen & R. R. Verma (Eds.), *The Philosophy of P. F. Strawson*. New Delhi: Indian Council of Philosophical Research.
- Clark, M. (2015). On the Rejection of Morality: Bernard Williams's Debt to Nietzsche. In *Nietzsche on Ethics and Politics* (pp. 41-61). Oxford: Oxford University Press.
- Clarke, R., & Rawling, P. (2023). Blameworthiness and Dependence. *The Philosophical Quarterly*, 74(1), 110-124. doi:10.1093/pq/pqad011
- Coates, D. J. (2017). Strawson's modest transcendental argument. *British Journal for the History of Philosophy*, 25(4), 799-822.
- Coates, D. J., & Tognazzini, N. A. (2013). The Contours of Blame. In D. J. Coates & N. A. Tognazzini (Eds.), *Blame: Its Nature and Norms* (pp. 3-26): Oxford University Press.
- Collingwood, R. G. (1940). *An Essay on Metaphysics*. London: Oxford University Press.
- Craig, E. (1990). *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Oxford: Clarendon Press.
- D'Arms, J., & Jacobson, D. (2000). The Moralistic Fallacy: On the 'Appropriateness' of Emotions. *Philosophical and Phenomenological Research*, 61(1), 65-90.
- D'Arms, J., & Jacobson, D. (2022). The Motivational Theory of Guilt (and Its Implications for Responsibility). In A. B. Carlsson (Ed.), *Self-Blame and Moral Responsibility* (pp. 11-27). Cambridge: Cambridge University Press.
- Darwall, S. (2006). *The Second-Person Standpoint: Morality, Respect, and Accountability*. Harvard University Press.
- Davidson, D. (1973-1974). On the Very Idea of a Conceptual Scheme. *Proceedings and Addresses of the American Philosophical Association*, 47, 5-20.
- De Mesel, B. (2018). Are our moral responsibility practices justified? Wittgenstein, Strawson and justification in 'Freedom and Resentment'. *British Journal for the History of Philosophy*, 26(3), 603-614.
- De Mesel, B. (2022a). Being and holding responsible: Reconciling the disputants through a meaning-based Strawsonian account. *Philosophical Studies*, 179(6), 1893-1913.

- De Mesel, B. (2022b). Ordinary Language Philosophy as an Extension of Ideal Language Philosophy. Comparing the Methods of the Later Wittgenstein and P.F. Strawson. *Philosophical Investigations*, 45(2), 175-199.
- De Mesel, B. (2022c). Taking the Straight Path. P.F. Strawson's Later Work on Freedom and Responsibility. *Philosophers' Imprint*, 22(12), 1-17.
- De Mesel, B. (forthcoming). Strawson, Skepticism and the Natural Roots of Responsibility. In J. Hyman & M. Thorne (Eds.), *Scepticism and Naturalism*. Leiden: Brill.
- De Mesel, B., & Cuypers, S. E. (2023). Strawson's Account of Morality and its Implications for Central Themes in 'Freedom and Resentment'. *Philosophical Quarterly*.
- De Mesel, B., & Heyndels, S. (2019). The Facts and Practices of Moral Responsibility. *Pacific Philosophical Quarterly*, 100(3), 790-811.
- Deigh, J. (2011). Reactive Attitudes Revisited. In C. Bagnoli (Ed.), *Morality and the Emotions*: Oxford University Press.
- Deonna, J. A., & Teroni, F. (2021). Which Attitudes for the Fitting Attitude Analysis of Value? *Theoria*, 87(5), 1099-1122. doi:<https://doi.org/10.1111/theo.12333>
- Ewing, A. C. (1947). *The Definition of Good*. London: Routledge & Kegan Paul.
- Ezcurdia, M. (1998). The concept-conception distinction. *Philosophical Issues*, 9, 187-192.
- Fischer, J. M. (2014). Peter Strawson and the Facts of Agency. In D. Shoemaker & N. A. Tognazzini (Eds.), *Oxford Studies in Agency and Responsibility: 'Freedom and Resentment' at 50* (Vol. 2): Oxford University Press.
- Fischer, J. M., & Ravizza, M. (1993). *Perspectives on Moral Responsibility*: Cornell University Press.
- Fischer, J. M., & Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*: Cambridge University Press.
- Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*: Oxford University Press.
- Fricker, M. (2016). What's the point of blame? A paradigm based explanation. *Noûs*, 50(1), 165-183.
- Garcia, A. G., & Green Werkmäster, J. (2018). Subjectivism and the Framework of Constitutive Grounds. *Ethical Theory and Moral Practice*, 21(1), 155-167.
- Gibbard, A. (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgement*. Oxford: Clarendon.
- Gibbon, E. (2004). *Memoirs of My Life and Writings* (P. J. Riddick & D. Widger Eds. 2012 ed.): Project Gutenberg.
- Glock, H.-J. (2003a). Introduction. In H.-J. Glock (Ed.), *Strawson and Kant* (pp. 1-5): Oxford University Press.

- Glock, H.-J. (2003b). Strawson and Analytic Kantianism. In H.-J. Glock (Ed.), *Strawson and Kant* (pp. 15-42): Oxford University Press.
- Glock, H.-J. (2007). Relativism, commensurability and translatability. *Ratio*, 20(4), 377-402.
- Glock, H.-J. (2009). Concepts, Conceptual Schemes and Grammar. *Philosophia*, 37, 653-668.
- Glock, H.-J. (2012). Strawson's descriptive metaphysics. In L. Haaparanta & H. Koskinen (Eds.), *Categories of Being: Essays on Metaphysics and Logic* (pp. 391): Oxford University Press, Usa.
- Glock, H.-J. (2016). Philosophy Rehinged? *International Journal for the Study of Skepticism*, 6(2-3), 274-308.
- Glock, H.-J. (2017). Ordinary Language Philosophy and Descriptive Metaphysics. In A. Preston (Ed.), *Analytic Philosophy: An Interpretative History* (pp. 214-228). New York, NY: Routledge.
- Glock, H.-J. (2022). Strawson and Non-Revisionary Naturalism. In M. D. Caro & D. Macarthur (Eds.), *The Routledge Handbook of Liberal Naturalism* (1st ed.): Routledge.
- Glock, H.-J. (2024). Concepts and Experience in *Bounds of Sense* and Beyond. In S. Heyndels, A. Bengtson, & B. De Mesel (Eds.), *P. F. Strawson and his Philosophical Legacy* (pp. 120-145): Oxford University Press.
- Gomes, A. (2016). Unity, Objectivity, and the Passivity of Experience. *European Journal of Philosophy*, 24(3), 946-969.
- Gomes, A. (2017). Perception and reflection. *Philosophical Perspectives*, 31(1), 131-152.
- Gomes, A. (2019, March 3). Snapshot: P. F. Strawson. *The Philosophers' Magazine*.
- Gomes, A. (2024). Strawson's Metacritique In S. Heyndels, A. Bengtson, & B. De Mesel (Eds.), *P.F. Strawson and His Legacy* (pp. 146-168): Oxford University Press.
- Grice, H. P., & Strawson, P. F. (1956). In Defense of a Dogma. *The Philosophical Review*, 65(2), 141-158. doi:10.2307/2182828
- Grundman, T., & Misselhorn, C. (2003). Transcendental Arguments and Realism. In H.-J. Glock (Ed.), *Strawson and Kant* (pp. 205-218): Oxford University Press.
- Hacker, P. M. S. (1996). *Wittgenstein's Place in Twentieth-Century Analytic Philosophy*: Blackwell Publishers Ltd.
- Hacker, P. M. S. (2001). On Strawson's Rehabilitation of Metaphysics. In *Wittgenstein: Connections and Controversies* (pp. 345-370): Oxford University Press.
- Hacker, P. M. S. (2019). *Wittgenstein: Meaning and mind*: Wiley-Blackwell.
- Hahn, L. E. (Ed.) (1998). *The Philosophy of P. F. Strawson*: Open Court.

- Hall, E. (2020). *Value, Conflict, and Order: Berlin, Hampshire, Williams, and the Realist Revival in Political Theory*. Chicago and London: The University of Chicago Press.
- Hampshire, S. (1983). *Morality and Conflict*. Oxford: Basil Blackwell.
- Hart, H. L. A. (2012). *The Concept of Law* (3rd ed.). Oxford: Oxford University Press.
- Hartmann, M. (2020). How a critical Humean naturalism is possible: Contesting the Neo-Aristotelian reading. *Philosophy and Social Criticism*.
- Hasan, A., & Fumerton, R. (2022). Foundationalist Theories of Epistemic Justification. *The Stanford Encyclopedia of Philosophy*. Retrieved from <https://plato.stanford.edu/entries/justep-foundational/>
- Heawood, J. (2006). Peter Strawson (1919-2006): A Sort of Obituary. *Philosophy Now*(57), 25-27.
- Helm, B. W. (2012). Accountability and some social dimensions of human agency. *Philosophical Issues*, 22(1), 217-232.
- Heyndels, S. (2019). Strawson's Method in 'Freedom and Resentment'. *Journal of Ethics*, 23(4), 407-423.
- Heyndels, S. (2020). *Philosophy as an Exercise in Conceptual Self-consciousness: An Explication of P.F. Strawson's Philosophical Methodology*. (Ph.D.). KU Leuven,
- Heyndels, S. (2022a). Excuses, Exemptions, and the Challenges to Social Naturalism. *International Journal of Philosophical Studies*, 1-14. doi:10.1080/09672559.2021.2022410
- Heyndels, S. (2022b). A forgotten logical expressivist: Strawson's philosophy of logic and its challenges. *Synthese*, 200(187), 1-23.
- Heyndels, S., Bengtson, A., & De Mesel, B. (2024a). Introduction. In S. Heyndels, A. Bengtson, & B. De Mesel (Eds.), *P. F. Strawson and his Philosophical Legacy* (pp. 1-14). Oxford: Oxford University Press.
- Heyndels, S., Bengtson, A., & De Mesel, B. (Eds.). (2024b). *P. F. Strawson and his Philosophical Legacy*. Oxford: Oxford University Press.
- Heyndels, S., & De Mesel, B. (2018). On Shoemaker's Response-Dependent Theory of Responsibility. *Dialectica*, 72(3), 445-451. Retrieved from <https://www.jstor.org/stable/48552660>
- Hieronymi, P. (2020). *Freedom, Resentment, and the Metaphysics of Morals*: Princeton University Press.
- Howard, C. (2018). Fittingness. *Philosophy Compass*, 13(11), e12542.
- Howard, C. (2023). Fitting Attitude Theories of Value. In E. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2023 ed.): Metaphysics Research Lab, Stanford University.
- Hume, D. (2011). *A Treatise of Human Nature*. In: Oxford University Press.

- Hutchison, K. (2018). Moral Responsibility, Respect, and Social Identity. In K. Hutchison, C. Mackenzie, & M. Oshana (Eds.), *Social Dimensions of Moral Responsibility*: Oxford University Press.
- Jefferson, A. (2019). Instrumentalism about Moral Responsibility Revisited. *Philosophical Quarterly*, 69(276), 555-573.
- Jubb, R. (2016). Recover it From the Facts as We Know Them. *Journal of Moral Philosophy*, 13(1), 77-99.
- Kant, I. (1781). Critique of Pure Reason. In.
- Krishnan, N. (2023). *A Terribly Serious Adventure: Philosophy at Oxford, 1900-1960*: Profile Books Ltd.
- Lacey, N. (2006). *A Life of H.L.A. Hart: The Nightmare and the Noble Dream*: Oxford University Press.
- Lalumera, E. (2014). On the explanatory value of the concept conception distinction. *Rivista Italiana di Filosofia del Linguaggio*, 8(2), 73-81.
- Macnamara, C. (2015a). Blame, Communication, and Morally Responsible Agency. In R. Clarke, M. McKenna, & A. Smith (Eds.), *The Nature of Moral Responsibility: New Essays* (pp. 211-236): Oxford University Press.
- Macnamara, C. (2015b). Reactive Attitudes as Communicative Entities. *Philosophy and Phenomenological Research*, 90(3), 546-569.
- Magee, B., & Strawson, P. F. (1971). Conversation with Peter Strawson. In B. Magee (Ed.), *Modern British Philosophy* (pp. 115-130). London: Secker & Warburg.
- Magee, B., & Williams, B. (1971). Conversation with Bernard Williams: Philosophy and Morals. In B. Magee (Ed.), *Modern British Philosophy* (pp. 150-165). London: Secker & Warburg.
- Martinisch, A. P. (2024). P. F. Strawson and Connective Analysis. In S. Heyndels, A. Bengtson, & B. De Mesel (Eds.), *P. F. Strawson and his Philosophical Legacy* (pp. 213-233): Oxford University Press.
- McDowell, J. (1985). Values and Secondary Qualities. In T. Honderich (Ed.), *Morality and Objectivity* (pp. 110-129). London: Routledge & Kegan Paul.
- McDowell, J. (1999). Virtue and Reason. In *Mind, Value, & Reality* (pp. 50-73): Harvard University Press.
- McGeer, V. (2012). Co-Reactive Attitudes and the Making of Moral Community. In R. Langdon & C. MacKenzie (Eds.), *Emotions, Imagination and Moral Reasoning* (pp. 299-326). New York: Psychology Press.
- McGeer, V. (2013). Civilizing blame. In D. J. Coates & N. A. Tognazzini (Eds.), *Blame: Its Nature and Norms* (pp. 162--188): Oxford University Press.
- McGeer, V. (2014). P.F. Strawson's Consequentialism. In D. Shoemaker & N. A. Tognazzini (Eds.), *Oxford Studies in Agency and Responsibility: 'Freedom and Resentment' at 50* (Vol. 2): Oxford University Press.

- McGeer, V. (2019). Scaffolding agency: A proleptic account of the reactive attitudes. *European Journal of Philosophy*, 27(2), 301-323.
- McHugh, C., & Way, J. (2016). Fittingness First. *Ethics*, 126(3), 575-606.
- McKenna, M. (2005). Where Frankfurt and Strawson meet. *Midwest Studies In Philosophy*, 29(1), 163-180.
- McKenna, M. (2008). The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism. In M. McKenna & P. Russell (Eds.), *Free Will and Reactive Attitudes* (pp. 201-218): Routledge. (Reprinted from: 2016).
- McKenna, M. (2012). *Conversation & Responsibility*. New York: Oxford University Press.
- McKenna, M. (2019a). Basically Deserved Blame and its Value. *Journal of Ethics and Social Philosophy*, 15(3).
- McKenna, M. (2019b). The Free Will Debate and Basic Desert. *The Journal of Ethics*, 23(3), 241-255.
- McKenna, M. (2019c). Further Reflections on The Free Will Debate and Basic Desert: A Reply to Nelkin and Pereboom. *The Journal of Ethics*, 23(3), 277-290.
- McKenna, M., & Pereboom, D. (2016). *Free Will: A Contemporary Introduction*: Routledge.
- McKenna, M., & Russell, P. (2008a). Perspectives on P.F. Strawson's "Freedom and Resentment". In M. McKenna & P. Russell (Eds.), *Free Will and Reactive Attitudes*: Routledge.
- McKenna, M., & Russell, P. (Eds.). (2008b). *Free Will and Reactive Attitudes*: Routledge.
- Menges, L. (2017). Grounding Responsibility in Appropriate Blame. *American Philosophical Quarterly*, 54(1), 15-24.
- Menges, L. (2020). Responsibility and appropriate blame: The no difference view. *European Journal of Philosophy*, 29(2), 393-409.
- Menges, L. (2023). Responsibility, Free Will, and the Concept of Basic Desert. *Philosophical Studies*, 180(2), 615-636.
- Milam, P.-E. (2016). Reactive Attitudes and Personal Relationships. *Canadian Journal of Philosophy*, 46(1), 102-122. doi:10.1080/00455091.2016.1146032
- Milam, P.-E. (2021). Get Smart: Outcomes, Influence, and Responsibility. *The Monist*, 104(4), 443-457.
- Miller, D. E. (2014a). "Freedom and Resentment" and Consequentialism. *Journal of Ethics and Social Philosophy*, 8(2), 1-23.
- Miller, D. E. (2014b). Reactive Attitudes and the Hare-Williams Debate: Towards a New Consequentialist Moral Psychology. *Philosophical Quarterly*, 64(254), 39-59.

- Moore, G. E. (1925). A Defence of Common Sense. In T. Baldwin (Ed.), *G. E. Moore: Selected Writings*: Routledge.
- Moore, G. E. (1939). Proof of an External World. In T. Baldwin (Ed.), *G. E. Moore: Selected Writings*: Routledge.
- Moyal-Sharrock, D. (2004a). On Certainty and the grammaticalization of experience. In D. Moyal-Sharrock (Ed.), *The Third Wittgenstein: The Post-Investigations Works* (pp. 63-78): Ashgate.
- Moyal-Sharrock, D. (2004b). *The Third Wittgenstein: The Post-Investigations Works*: Ashgate.
- Murdoch, I. (2001). *The Sovereignty of Good*: Routledge.
- Naar, H. (2021). The fittingness of emotions. *Synthese*, 199(5-6), 13601-13619.
- Nagel, T. (1986). *The View from Nowhere*. New York: Oxford University Press.
- Nelkin, D. K. (2019). Desert, Free Will, and Our Moral Responsibility Practices. *The Journal of Ethics*, 23(3), 265-275.
- Nietzsche, F. (1889/1997). *Twilight of the Idols; Or, How to Philosophize with the Hammer* (R. Polt, Trans.). Indianapolis: Hackett.
- Nietzsche, F. (1986). *Human, All Too Human* (R. J. Hillingdale, Trans.): Cambridge University Press.
- Nietzsche, F. (2003). *The Genealogy of Morals* (H. B. Samuel, Trans. T. N. R. Rogers Ed.): Dover Publishing.
- Nietzsche, F. (2014). Beyond Good and Evil: Prelude to a Philosophy of the Future (R. J. Hollingdale, Trans.). In: Penguin Classics.
- Nowell-Smith, P. H. (1948). Freewill and moral responsibility. *Mind*, 57(225), 45-61.
- O'Grady, J. (2006). Sir Peter Strawson Obituary. *The Guardian*.
- Pears, D. (1998). Strawson on Freedom and Resentment. In L. E. Hahn (Ed.), *The Philosophy of P. F. Strawson* (pp. 244-258): Open Court.
- Pereboom, D. (1995). Determinism *al Dente*. *Noûs*, 29, 21-45.
- Pereboom, D. (2001). *Living Without Free Will*: Cambridge University Press.
- Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life*: Oxford University Press.
- Pereboom, D. (2016). Transcendental Arguments. In H. Cappelen, T. S. Gendler, & J. Hawthorne (Eds.), *The Oxford Handbook of Philosophical Methodology*: Oxford University Press.
- Pereboom, D. (2019). What Makes the Free Will Debate Substantive? *The Journal of Ethics*, 23(3), 257-264.
- Pereboom, D. (2021a). Undivided Forward-Looking Moral Responsibility. *The Monist*, 104(4), 484-497.
- Pereboom, D. (2021b). *Wrongdoing and the Moral Emotions*: Oxford University Press.

- Pereboom, D. (2022). Kant's Transcendental Arguments. In E. N. Zalta & U. Nobelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2022 ed.).
- Pettit, P. (2006/7). The Law of Demand. *Times Literary Supplement*(5422), 27.
- Pivčević, E. (1989). An Interview with Professor Sir Peter Strawson. *Cogito*, 3(1), 3-8.
- Prasad, R. (1995). Reactive Attitudes, Rationality and Determinism. In P. K. Sen & R. R. Verma (Eds.), *The Philosophy of P. F. Strawson* (pp. 346-376). New Delhi: Indian Council of Philosophical Research.
- Putnam, H. (1995). Logical Positivism, the Kantian Tradition and the Bounds of Sense. In P. K. Sen & R. R. Verma (Eds.), *The Philosophy of P. F. Strawson*. New Delhi: Indian Council of Philosophical Research.
- Putnam, H. (1998). Strawson and Skepticism. In L. E. Hahn (Ed.), *The Philosophy of P. F. Strawson* (pp. 273-287): Open Court.
- Queloz, M. (2018). Williams's Pragmatic Genealogy and Self-Effacing Functionality. *Philosophers' Imprint*, 18, 1-20.
- Queloz, M. (2020). How genealogies can affect the space of reasons. *Synthese*, 197(5), 2005-2027.
- Queloz, M. (2021a). *The Practical Origins of Ideas: Genealogy as Conceptual Reverse-Engineering*: Oxford University Press.
- Queloz, M. (2021b). The Self-Effacing Functionality of Blame. *Philosophical Studies*, 178, 1361-1379. doi:<https://doi.org/10.1007/s11098-020-01479-y>
- Queloz, M. (2022a). The Essential Superficiality of the Voluntary and the Moralization of Psychology. *Philosophical Studies*, 179, 1591-1516-1520. doi:<https://doi.org/10.1007/s11098-021-01720-2>
- Queloz, M. (2022b). A Shelter from Luck: The Morality System Reconstructed. In A. Szigeti & M. Talbert (Eds.), *Morality and Agency: Themes from Bernard Williams*: Oxford University Press.
- Queloz, M. (2023). The Dworkin–Williams debate: Liberty, conceptual integrity, and tragic conflict in politics. *Philosophy and Phenomenological Research*, n/a(n/a). doi:<https://doi.org/10.1111/phpr.13002>
- Queloz, M., & Cueni, D. (2019). Nietzsche as a Critic of Genealogical Debunking: Making Room for Naturalism without Subversion. *The Monist*, 102(3), 277-297.
- Queloz, M., & Cueni, D. (2021). Left Wittgensteinianism. *European Journal of Philosophy*, 29, 758-777. doi:<https://doi.org/10.1111/ejop.12603>
- Quine, W. V. O. (1951). Main Trends in Recent Philosophy: Two Dogmas of Empiricism. *The Philosophical Review*, 60(1), 20-43. doi:10.2307/2181906
- Quine, W. V. O. (1969). *Ontological Relativity*: Columbia University Press.
- Rabinowicz, W., & Rønnow-Rasmussen, T. (2004). The strike of the demon: On fitting pro-attitudes and value. *Ethics*, 114(3), 391-423.

- Rawls, J. (1971). *A Theory of Justice* (Reprint, 50th Anniversary ed.): Harvard University Press.
- Rorty, R. (1971). Verificationism and transcendental arguments. *Noûs*, 5(1), 3-14.
- Rorty, R. (Ed.) (1992). *The Linguistic Turn: Essays in Philosophical Method: With Two Retrospective Essays*. Chicago: The University of Chicago Press.
- Rummens, S., & De Mesel, B. (2023). A Wittgensteinian Account of Free Will and Moral Responsibility. In C. Eriksen, J. Hermann, N. O'Hara, & N. Pleasants (Eds.), *Philosophical Perspectives on Moral Certainty* (pp. 132-155): Routledge.
- Russell, P. (1995). *Freedom and Moral Sentiment: Hume's Way of Naturalizing Responsibility*. New York: Oxford University Press.
- Russell, P. (2017a). Free Will and Moral Sentiments: Strawsonian Theories. In K. Timpe, M. Griffith, & N. Levy (Eds.), *The Routledge Companion to Free Will*. London: Routledge.
- Russell, P. (2017b). Moral Sense and the Foundations of Responsibility In P. Russell (Ed.), *The Limits of Free Will*: Oxford University Press.
- Russell, P. (2017c). Responsibility, Naturalism, and "The Morality System". In P. Russell (Ed.), *The Limits of Free Will* (pp. 97-118). New York: Oxford University Press.
- Russell, P. (2017d). Strawson's Way of Naturalizing Responsibility. In P. Russell (Ed.), *The Limits of Free Will*. New York: Oxford University Press.
- Russell, P. (2021). Responsibility Skepticism and Strawson's Naturalism: Review Essay on Pamela Hieronymi, *Freedom, Resentment & The Metaphysics of Morals* (Princeton: Princeton University Press, 2020). *Ethics*, 131(4), 754-776.
- Russell, P. (2024). Responsibility after 'Morality': Strawson's Naturalism and Williams' Genealogy. In S. Heyndels, A. Bengtson, & B. De Mesel (Eds.), *P. F. Strawson and His Philosophical Legacy*. Oxford: Oxford University Press.
- Salmon, W. (1957). Should We Attempt to Justify Induction? *Philosophical Studies*(8), 33-40.
- Sars, N. (2022a). Incapacity, Inconceivability, and Two Types of Objectivity. *Pacific Philosophical Quarterly*, 103(1), 76-94.
- Sars, N. (2022b). Strawson's underappreciated argumentative structure. *European Journal of Philosophy*.
- Scanlon, T. M. (1988). The Significance of Choice. In S. M. McMurrin (Ed.), *The Tanner Lectures on Human Values* (Vol. 8, pp. 149-216). Salt Lake City: University of Utah Press.
- Scanlon, T. M. (1998). *What we owe to each other* (paperback edition, 2000 ed.): Harvard University Press.
- Scanlon, T. M. (2003). The Significance of Choice. In G. Watson (Ed.), *Free Will* (2nd ed., pp. 352-371): Oxford University Press.

- Schlick, M. (1939). *The Problems of Ethics* (D. Rynin, Trans.). New York: Prentice Hall.
- Sellars, W. (1997). *Empiricism and the Philosophy of Mind*. Cambridge, Massachusetts: Harvard University Press.
- Sen, P. K., & Verma, R. R. (1995). *The Philosophy of P. F. Strawson*. New Delhi: Indian Council of Philosophical Research.
- Shabo, S. (2012). Where Love and Resentment Meet: Strawson's Intrapersonal Defense of Compatibilism. *Philosophical Review*, 121(1), 95-124.
- Shabo, S. (2022). Responsibility, Personal Relationships, and the Significance of the Reactive Attitudes. In D. K. Nelkin & D. Pereboom (Eds.), *The Oxford Handbook of Moral Responsibility*: Oxford University Press.
- Shoemaker, D. (2007). Moral Address, Moral Responsibility, and the Boundaries of the Moral Community. *Ethics*, 118(1), 70-108.
- Shoemaker, D. (2015). *Responsibility From the Margins*: Oxford University Press.
- Shoemaker, D. (2017). Response-dependent responsibility; or, a funny thing happened on the way to blame. *Philosophical Review*, 126(4), 481-527.
- Shoemaker, D. (2020). Responsibility: the State of the Question Fault Lines in the Foundations. *Southern Journal of Philosophy*, 58(2), 205-237.
- Shoemaker, D. (2022). Response-Dependent Theories of Responsibility. In D. K. Nelkin & D. Pereboom (Eds.), *The Oxford Handbook of Moral Responsibility*: Oxford University Press.
- Shoemaker, D., & Tognazzini, N. (Eds.). (2014). *Oxford Studies in Agency and Responsibility: 'Freedom and Resentment' at 50* (Vol. 2). New York: Oxford University Press.
- Shoemaker, D., & Vargas, M. (2021). Moral Torch Fishing: A signaling theory of blame. *Noûs*.
- Shoemaker, S. (1963). *Self-Knowledge and Self-Identity*. Ithaca, N.Y.: Cornell University Press.
- Sidgwich, H. (1874). The Method of Ethics. In. Cambridge: Cambridge University Press. (Reprinted from: 2011).
- Sie, M., & Pereboom, D. (2015). *Basic Desert, Reactive Attitudes and Free Will*: Routledge.
- Sliwa, P. (2019). Reverse-engineering blame. *Philosophical Perspectives*, 33(1), 200-219.
- Smart, J. J. C. (1961). Free will, Praise and Blame In G. Watson (Ed.), *Free will* (2nd ed., pp. 58-71). Oxford, UK: Oxford University Press. (Reprinted from: 2003).
- Smilansky, S. (2000). *Free will and Illusion*: Oxford University Press.

- Smilansky, S. (2008). Free Will: From Nature to Illusion. In M. McKenna & P. Russell (Eds.), *Free Will and Reactive Attitudes* (pp. 235-253): Routledge. (Reprinted from: 2016).
- Smith, A. (2012). Moral Blame and Moral Protest. In D. J. Coates & N. A. Tognazzini (Eds.), *Blame: Its Nature and Norms* (pp. 27-48): Oxford University Press.
- Snowdon, P., & Gomes, A. (2023). Peter Frederick Strawson. In E. N. Zalta & U. Nobelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2023 ed.).
- Sommers, T. (2007). The objective attitude. *Philosophical Quarterly*, 57(228), 321–341.
- Sommers, T. (2012). *Relative Justice: Cultural Diversity, Free Will, and Moral Responsibility*: Princeton University Press.
- Srinivasan, A. (2015). The Archimedean Urge. *Philosophical Perspectives*, 29(1), 325-362.
- Stern, R. (2003). On Strawson's Naturalistic Turn. In H.-J. Glock (Ed.), *Strawson and Kant*. Oxford: Oxford University Press.
- Stern, R. (2019). Transcendental Arguments. *The Stanford Encyclopedia of Philosophy*. Fall 2021. Retrieved from <https://plato.stanford.edu/archives/fall2021/entries/transcendental-arguments/>
- Strawson, P. F. (1950). On Referring. *Mind*, 59(235), 320-344.
- Strawson, P. F. (1950). Truth. *Proceedings of the Aristotelian Society*, 24, 129-156.
- Strawson, P. F. (1952). *Introduction to Logical Theory*: Methuen.
- Strawson, P. F. (1958). On Justifying Induction. *Philosophical Studies*(9), 20-21.
- Strawson, P. F. (1959). *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen: Univeristy Paperbacks.
- Strawson, P. F. (1961). Social Morality and Individual Ideal. *Philosophy*, 36(136), 1 - 17.
- Strawson, P. F. (1963). Carnap's Views on Conceptual Systems versus Natural Languages in Analytic Philosophy. In P. A. Schilpp (Ed.), *The Philosophy of Rudolf Carnap* (pp. 503-518): Open Court: La Salle.
- Strawson, P. F. (1966). *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason* (e-book 2005 ed.): Routledge.
- Strawson, P. F. (1968). Bennett on Kant's Analytic. *The Philosophical Review*, 77(3), 332-339.
- Strawson, P. F. (1971). *Logico-Linguistic Papers*. London: Methuen.
- Strawson, P. F. (1974). *Subject and Predicate in Logic and Grammar*. London: Methuen.
- Strawson, P. F. (1980). P. F Strawson Replies. In Z. van Straaten (Ed.), *Philosophical Subjects* (pp. 260-296): Clarendon Press.

- Strawson, P. F. (1982). Transcendental Arguments and Science by Peter Bieri, Rolf-Peter Horstmann, and Lorenz Kruger. *Journal of Philosophy*, 79(1), 45-50.
- Strawson, P. F. (1985). *Analyse et métaphysique*. Paris: Vrin.
- Strawson, P. F. (1992a). *Analysis and Metaphysics: An Introduction to Philosophy*. New York: Oxford University Press.
- Strawson, P. F. (1992b). Freedom and Necessity. In *Analysis and Metaphysics: An Introduction to Philosophy* (pp. 133-142). New York: Oxford University Press.
- Strawson, P. F. (1992c, August 9th). [Letter to Paul Russell].
- Strawson, P. F. (1995). Replies. In P. K. Sen & R. R. Verma (Eds.), *The Philosophy of P. F. Strawson* (pp. 398-433). New Delhi: Indian Council of Philosophical Research.
- Strawson, P. F. (1997). *Entity and Identity and Other Essays*. Oxford: Oxford University Press.
- Strawson, P. F. (1998a). Intellectual Autobiography. In L. E. Hahn (Ed.), *The Philosophy of P. F. Strawson*: Open Court. (Reprinted from: *Scepticism and Naturalism* (2008)).
- Strawson, P. F. (1998b). Reply to David Pears. In L. E. Hahn (Ed.), *The Philosophy of P. F. Strawson*: Open Court.
- Strawson, P. F. (1998c). Reply to Ernest Sosa. In L. E. Hahn (Ed.), *The Philosophy of P. F. Strawson* (pp. 370-372): Open Court.
- Strawson, P. F. (1998d). Reply to Hilary Putnam. In L. E. Hahn (Ed.), *The Philosophy of P. F. Strawson* (pp. 288-292): Open Court.
- Strawson, P. F. (1998e). Reply to Simon Blackburn. In L. E. Hahn (Ed.), *The Philosophy of P. F. Strawson*: Open Court.
- Strawson, P. F. (2003). A Bit of Intellectual Autobiography. In H.-J. Glock (Ed.), *Strawson and Kant* (pp. 6-14): Oxford University Press.
- Strawson, P. F. (2005). Preface. In *Scepticism and Naturalism*: Taylor & Francis E-library.
- Strawson, P. F. (2008a). Freedom and Resentment. In *Freedom and Resentment and other essays*: Routledge.
- Strawson, P. F. (2008b). *Freedom and Resentment and other essays* (Foreward by Paul Snowden ed.). Abingdon: Routledge.
- Strawson, P. F. (2008c). *Scepticism and Naturalism: Some Varieties*: Routledge.
- Strawson, P. F. (2008d). Social Morality and Individual Ideal. In P. F. Strawson (Ed.), *Freedom and Resentment and Other Essays*. London: Routledge.
- Strawson, P. F. (2008e). Wittgenstein's Philosophical Investigations. In *Freedom and Resentment and other essays* (pp. 147-185): Routledge.
- Strawson, P. F. (2011a). Analysis, Science, and Metaphysics. In P. F. Strawson (Ed.), *Philosophical Writings*. Oxford: Oxford University Press.

- Strawson, P. F. (2011b). Construction and Analysis. In P. F. Strawson (Ed.), *Philosophical Writings*. Oxford: Oxford University Press.
- Strawson, P. F. (2011c). Liberty and Necessity. In P. F. Strawson (Ed.), *Philosophical Writings*: Oxford University Press.
- Strawson, P. F. (2011d). *Philosophical Writings* (G. Strawson & M. Montague Eds.).
- Strawson, P. F. (2011e). Why Philosophy? In *Philosophical Writings*. Oxford: Oxford University Press.
- Strawson, P. F. (2019). *The Bounds of Sense*: Routledge.
- Strawson, P. F., Ezcurdia, M., Sainsbury, M., & Davies, M. (2008). Sino-British Summer School, 1992 (Interview Typescript). In R. Walker (Ed.), *A Tribute to Sir Peter Strawson* (Vol. 7, pp. 75-114): Magdalen College Oxford, Alden Group.
- Stroud, B. (1968). Transcendental arguments. *Journal of Philosophy*, 65(9), 241-256.
- Stroud, B. (1979). The Significance of Scepticism. In P. Bieri, R.-P. Horstmann, & L. Krüger (Eds.), *Transcendental Arguments and Science*: Reidel.
- Stroud, B. (2002). The Goal of Transcendental Arguments. In B. Stroud (Ed.), *Understanding Human Knowledge: Philosophical Essays* (pp. 203-223): Oxford University Press.
- Suikkanen, J. (2009). Buck-Passing Accounts of Value. *Philosophy Compass*, 4, 769-779.
- Szigeti, A. (2012). Revisiting Strawsonian Arguments from Inescapability. *Philosophica*, 85(2), 91-121.
- Talbert, M. (2008). Blame and responsiveness to moral reasons: Are psychopaths blameworthy? *Pacific Philosophical Quarterly*, 89(4), 516-535.
- Talbert, M. (2012). Moral Competence, Moral Blame, and Protest. *The Journal of Ethics*, 16(1), 89-109.
- Thomson, J. J. (1964). Private Languages. *American Philosophical Quarterly*, 1(1), 20-31.
- Todd, P. (2016). Strawson, Moral Responsibility, and the "Order of Explanation": An Intervention. *Ethics*, 127(1), 208-240.
- Todd, P. (manuscript). Strawsonian Moral Responsibility, Response-dependence, and the Possibility of Global Error. 1-29.
- Tognazzini, N. A. (2013). Blameworthiness and the Affective Account of Blame. *Philosophia*, 41(4), 1299-1312. doi:10.1007/s11406-013-9428-3
- Vanrie, W. (2021). What We All Know: Community in Moore's "A Defence of Common Sense". *Journal of the History of Philosophy*, 59(4), 629-651.
- Vargas, M. (2004). Responsibility and the aims of theory: Strawson and revisionism. *Pacific Philosophical Quarterly*, 85(2), 218-241.

- Vargas, M. (2013). *Building Better Beings: A Theory of Moral Responsibility*: Oxford University Press.
- Vargas, M. (2022). Instrumentalist Theories of Moral Responsibility. In D. K. Nelkin & D. Pereboom (Eds.), *The Oxford Handbook of Moral Responsibility*: Oxford University Press.
- Walker, R. (1989). Transcendental Arguments and Scepticism. In *Reading Kant* (pp. 55-76).
- Walker, R. (2008). *A Tribute to Sir Peter Strawson* (Vol. 7): Magdalen College Oxford, Alden Group.
- Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*: Harvard University Press.
- Watson, G. (2004). Responsibility and the Limits of Evil: Variations on a Strawsonian theme. In G. Watson (Ed.), *Agency and Answerability* (pp. 219-259). New York: Oxford University Press.
- Watson, G. (2008). Responsibility and the Limits of Evil: Variations on a Strawsonian theme. In M. McKenna & P. Russell (Eds.), *Free Will and Reactive Attitudes* (pp. 115-141): Oxford University Press.
- Watson, G. (2011). The Trouble with Psychopaths. In R. J. Wallace, R. Kumar, & S. Freeman (Eds.), *Reasons and recognition: Essays on the philosophy of TM Scanlon*: Oxford University Press.
- Watson, G. (2014). Peter Strawson on Responsibility and Sociality. In D. Shoemaker & N. A. Tognazzini (Eds.), *Oxford studies in agency and responsibility* (Vol. 2, pp. 15-32): Oxford University Press.
- Werkmäster, M. J. (2023). *Aspects of Blame: In which the nature of blame, blameworthiness, the standing to blame and proportional blame are discussed*. (PhD). Lund University, Lund.
- Williams, B. (1972). *Morality: An Introduction to Ethics* (Canto edition 1993 ed.). Cambridge, UK: Cambridge University Press.
- Williams, B. (1973). Morality and the Emotions. In *Problems of the Self: Philosophical Papers 1956-1972* (pp. 207-229). Cambridge, UK: Cambridge University Press.
- Williams, B. (1985). *Ethics and the Limits of Philosophy* (Routledge Classics 2011 ed.): Routledge.
- Williams, B. (1993). *Shame and Necessity*. London: Univeristy of California Press.
- Williams, B. (1995a). How free does the will need to be? In *Making Sense of Humanity and other philosophical papers, 1983-1993* (pp. 3-21). Cambridge: Cambridge University Press.
- Williams, B. (1995b). Internal reasons and the obscurity of blame. In *Making Sense of Humanity and other philosophical papers, 1983-1993*: Cambridge University Press.

- Williams, B. (1995c). The point of view of the universe: Sidgwick and the ambitions of ethics. In B. Williams (Ed.), *Making Sense of Humanity and other philosophical papers, 1983-1993* (pp. 153-171). Cambridge: Cambridge University Press.
- Williams, B. (1995d). Replies. In J. E. J. Altham & R. Harrison (Eds.), *World, Mind, and Ethics: Essays on the ethical philosophy of Bernard Williams* (pp. 185-224): Cambridge University Press.
- Williams, B. (2002). *Truth and Truthfulness: An Essay in Genealogy* (Vol. 78): Princeton University Press.
- Williams, B. (2006a). An Essay on Collingwood. In M. Burnyeat (Ed.), *The Sense of the Past: Essays in the History of Philosophy* (pp. 341-358): Princeton University Press.
- Williams, B. (2006b). Philosophy as a Humanistic Discipline. In A. W. Moore (Ed.), *Philosophy as a Humanistic Discipline*. Princeton, New Jersey: Princeton University Press.
- Williams, B. (2014). Why Philosophy Needs History. In M. Woods (Ed.), *Essays and Reviews, 1959-2002* (pp. 405-412). Princeton: Princeton University Press.
- Williams, M. (2005). Why Wittgenstein isn't a foundationalist. In D. Moyal-Sharrock & W. H. Brenner (Eds.), *Readings of Wittgenstein's On Certainty* (pp. 47-58). London: Palgrave Macmillan UK.
- Wittgenstein, L. (1969). *On Certainty* (G. E. M. Anscombe & G. H. von Wright Eds. Harper Torchbooks 1972 ed.): Basil Blackwell.
- Wittgenstein, L. (1978). *Remarks on the foundations of mathematics*: Blackwell.
- Wittgenstein, L. (2009a). *Philosophical Investigations* (G. E. M. Anscombe, P. M. S. Hacker, & J. Schulte, Trans. P. M. S. Hacker & J. Schulte Eds. rev. 4th ed.). Chichester, U.K.: Wiley-Blackwell.
- Wittgenstein, L. (2009b). Philosophy of Psychology—A Fragment (G. E. M. Anscombe, P. M. S. Hacker, & J. Schulte, Trans.). In P. M. S. Hacker & J. Schulte (Eds.), *Philosophical Investigations*: Wiley-Blackwell.
- Wolf, S. (1981). The Importance of Free Will. *Mind*, xc, 386-405.
- Yao, V. (2023). The Good Fit. *Philosophy and Phenomenological Research*.

